

must be reproduced?

(7M)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distributed under license by Oxford Molecular Ltd

DISTRIBUTION FIGURES BY ORIGIN MOLECULAR ETC

Distribution figures by Oxford molecular bio

```
Run on:      Wed Apr 19 23:38:47 2000;      MasPar time 17.05 Seconds
Tabular output not generated.              860.240 Million cell updates/sec
```

tabular output not generated:

Title: >US-09-323-597-2

Description: (1-491) from US0932359/.pdf

Perfect Score: 36/3

Sequence: 1 MALNSGSPPAIGPYENHG...VYGNVMVF¹LDNLYKQMKADG 491

Scoring table: PAM 150

Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database: swiss-prot38

1:sw1sprot

Statistics: Mean 48.170; Variance 69.227; scale 0.696

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	3635	99.0	452	1	TMS2_HUMAN	TRANSMEMBRANE PROTEASE	0.00e+00
2	895	24.4	1035	1	ENTK_BOVIN	ENTEROPEPTIDASE PRECUR	1.37e-2
3	890	24.2	1019	1	ENTK_HUMAN	ENTEROPEPTIDASE PRECUR	3.79e-22
4	862	23.5	1034	1	ENTK_PIG	ENTEROPEPTIDASE PRECUR	3.79e-22
5	845	23.0	1069	1	ENTK_MOUSE	ENTEROPEPTIDASE (EC 3.	2.73e-11
6	762	20.7	658	1	KAL_HUMAN	PLASMA KALLIKREIN PREC	1.15e-11
7	740	20.1	658	1	KAL_RAT	PLASMA KALLIKREIN PREC	1.97e-11
8	720	19.6	658	1	KAL_MOUSE	PLASMA KALLIKREIN PREC	8.98e-11
9	694	18.9	635	1	FA1L_HUMAN	COAGULATION FACTOR XI	1.96e-11
10	680	18.5	417	1	HEPS_HUMAN	SERINE PROTEASE HESPIN	1.72e-11
11	660	18.0	416	1	HEPS_MOUSE	SERINE PROTEASE HESPIN	7.22e-11
12	658	17.9	902	1	EPIT_MOUSE	EPITRIN (EC 3.4.21.-)	2.63e-11
13	652	17.8	416	1	HEPS_RAT	SERINE PROTEASE HESPIN	1.27e-11
14	634	17.3	812	1	PLMN_MOUSE	PLASMINOGEN PRECURSOR	1.41e-11
15	613	16.8	358	1	PLMN_HORSE	PLASMINOGEN (EC 3.4.21	8.01e-11
16	613	16.7	750	1	PLMN_PIG	PLASMINOGEN (EC 3.4.21	1.05e-11
17	608	16.6	810	1	PLMN_WACU	PLASMINOGEN PRECURSOR	2.61e-11
18	607	16.5	270	1	TRY1_MERCU	MAST CELL TRYPTASE PRE	4.96e-11
19	597	16.3	810	1	PLMN_HUMAN	PLASMINOGEN PRECURSOR	5.78e-11
20	596	16.2	343	1	PLMN_SHEEP	PLASMINOGEN (EC 3.4.21	3.04e-11
21	587	16.0	273	1	MCT7_RAT	MAST CELL PROTEASE 7 P	1.85e-11
22	586	16.0	363	1	PLMN_CANFA	PLASMINOGEN (EC 3.4.21	3.15e-11
23	583	15.9	267	1	TRY1_ANOGA	TRYPSIN 7 PRECURSOR (E	2.40e-11

45	538	14.6	263	1	CYR1_GADMO	CHYMOTRYPSIN A PRECURS	7.04e-111
44	540	14.7	277	1	TRY2_ANOGA	TRYPsin 2 PRECURSOR (E	1.97e-112)
43	544	14.8	436	1	ACRO_MOUSE	ACROSIN PRECURSOR (EC	1.55e-112)
42	546	14.9	675	1	TRY1_ANOGA	TRYPsin 1 PRECURSOR (E	1.35e-112)
41	545	14.9	274	1	BOFA_HUMAN	HEPATOCYTE GROWTH FACT	4.34e-113
40	547	14.9	415	1	ACRO_PIG	ACROSIN PRECURSOR (EC	2.30e-113)
39	548	14.9	275	1	TRY4_ANOGA	TRYPsin 4 PRECURSOR (E	1.32e-113)
38	546	14.9	264	1	CYR1_HUMAN	CHYMOTRYPSIN-LIKE PROT	4.34e-113
37	551	15.0	275	1	TRY1_HUMAN	ALPHA-TRYPTASE PRECURS	4.34e-113
36	554	15.1	263	1	CYR1_HUMAN	CHYMOTRYPSIN B PRE	2.66e-115
35	555	15.1	245	1	TRY3_ANOGA	TRYPsin 3 PRECURSOR (E	1.41e-115)
34	557	15.2	275	1	TRY3_ANOGA	TRYPsin 3 PRECURSOR (E	1.93e-116)
33	558	15.2	245	1	TRY4_BOVIN	CHYMOTRYPSIN A (EC	2.08e-116)
32	561	15.3	274	1	TRYM_RAT	MAST CELL TRYPTASE PRE	3.07e-117
31	564	15.4	343	1	PSS8_HUMAN	PROSTASIN PRECURSOR (E	4.52e-118)
30	566	15.4	263	1	CYR2_CAVFA	CHYMOTRYPSINEN 2 PRE	4.52e-118
29	569	15.5	812	1	PLMN_BOVIN	PLASMINOGEN PRECURSOR	1.86e-119
28	570	15.5	810	1	PLMN_ERIEU	PLASMINOGEN PRECURSOR	9.79e-120
27	568	15.5	421	1	ACRO_HUMAN	ACROSIN PRECURSOR (EC	5.75e-119
26	572	15.6	275	1	TRYB_HUMAN	BETA-TRYPTASE PRECURSO	2.73e-120
25	577	15.7	276	1	MCT7_MOUSE	MAST CELL PROTEASE 6 P	1.12e-121
24	582	15.8	273	1	ACRO_MOUSE	ACROSIN PRECURSOR (EC	4.55e-121)

ALIGNMENTS

ID	RESULT	1	STANDARD:	PRT:	492 AA.
AC	015393:				
DR	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).				
GN	TPRSS2.				
OS	Homo sapiens (Human).				
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
CC	Eutheria; Primates; Carnivora; Homiidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 97468144.				
RA	PADOUNI-GIACOBINO A., CHEN H., PETTSCH M.C., ROSSIER C.,				
RA	ANTONARAKIS S.E.;				
RT	"Cloning of the TPRS22 gene, which encodes a novel serine protease				
RT	with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.;"				
RL	Genomics 44:309-320(1997).				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND				
CC	WEAKLY IN SEVERAL OTHER TISSUES.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY.				
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb.ch).				
CC	-----				
DR	EMBL; U75329; AAC51784.1; -				
DR	HSSP; P00763; IDPO.				
DR	MIM; 602060; -				
DR	PROSITE; PS00134; TRYPSIN_HIS. 1.				
DR	PROSITE; PS00135; TRYPSIN_SER. 1.				
DR	PROSITE; PS01209; LDLRA_1; 1.				
DR	PROSITE; PS50068; LDLRA_2; 1.				
DR	PFAM; PF00057; LDL_recept_a; 1.				
DR	PFAM; PF00089; trypsin; 1.				
KW	Hydrolase; serine protease; Transmembrane; Signal-anchor.				
FT	DOMAIN 1 84				
FT	TRANSMEM 85 105				
FT	POTENTIAL.				

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 2479)
Paoloni-Giacobino, A., Chen, H., Peltsch, M. C., Roessler, C. and Antonarakis, S. E.
TITLE Cloning of the TM6RS2 gene, which encodes a novel serine protease with transmembrane, IDRA, and SRCR domains and maps to 21q22.3
JOURNAL Genomics 44 (3), 309-320 (1997)
MEDLINE 97468144
REFERENCE 2 (bases 1 to 2479)
Paoloni-Giacobino, A., Chen, H. and Antonarakis, S. E.
AUTHORS Direct Submission
TITLE Submitted (17-OCT-1996) Medical Genetics, University of Geneva
JOURNAL Medical School, 1 Michel-Servet Street, Geneva 1211, Switzerland
FEATURES
source
1. 2479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
57..1535
/codon_start=1
/product="serine protease"
/protein_id="AAC51784.1"
/db_xref="GI:2507613"
/translation="MALNSGSPPAIPYENHGYDPENPYPAOPTVVPYVYENHYPAY
YSPVPOVAVRLTQASNPVYQPSPTVCTSTKALCTTLTFLVGAALAA
GLAKFNGSKNSGIECDSPGICNPSCDGEDNRCVRLGPNFIILO
YSSORLSMHPVCCDDNNENYRAACRDMKNNFYSSOGIVDSGSTFPMKLTSGN
VDYKRLHSDASSKAVYSLRCLAGVSNRSRIRYSGEALLEGAMWOVSLHO
NNAVCGSITTPETPIYTAACHVCKPLNNPHHTAPRAGILROSFPAGAGQONKYSK
PNYDSGTAKNDILMKLQKRLFTNDLKPCLPCLPMDLOPELCLISGATEEKG
TSEVLAARAKVLIETGRNSRYVDNLITPAMICAGFLQGNVDSGSGPLVTSNN
NIWMLIGTSMGSCAKARPYGVYGVNMYETDIIYROMKANG"
BASE COUNT 578 a 650 c 677 g 574 t
ORIGIN
Query Match 95.28; Score 1655.2; DB 11; Length 2479;
Best Local Similarity 99.58; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

|||||
Db 421 CCTACAGTACCTCATCAACACCCCTTAACGTGATAGCGGTGCACACTGCCGCGG 480
Qy 536 gggaagagagaaatggtgtgttcgctctcaagaccacactcatccttcagggttact 595
Db 481 GGGAGGACGAATCGGTGTGGCTCTACCGACCAACCTTCACTTCAGATTAAT 540
Qy 596 catctcaagaaatctctgacccctgtgtgccaagaagcactggaacgaaactaagggc 655
Db 541 CATCTCAGAGGAAGTCTGTGGCACCTCTGTGCGCAACAGACTGGAAGAACTACGGCG 600
Qy 656 gggcgagcctgcaaggacatggtgctataagaataatttctactagccaaggaatagtg 715
Db 601 GGGGCGCCTCAGAGACATGGCTATAGAAATATTTTACTATACCAAGGAATAGTGG 660
Qy 716 atgacagcgatcccaacagcttatagaactgaaacaaagtcggcgcaatgtcatatc 775
Db 661 ATGACACGGATCCACAGCTTATGAACTGAACAACTGCGGCAATGTGAAATCT 720
Qy 776 ataaaaactgtacacacagtgatgacctgtctctcaaaagcagtggttcttaccgtga 835
Db 721 ATAAAAACTGTACACAGTGAATGCTGTTCTTCAAAAGCAGTGTTCCTTACGCTGTT 780
Qy 836 taagctgcyggtcaacttgaactcaagccgcagagcaggaattgtgcygcgagagcg 895
Db 781 TAGCCTCGGGGTCACCTTGAATCAAGCCGACAGACAGATCGTGGCGGTGAGAGCG 840
Qy 896 cgtcccgagggtccctgacctgcaaggtcaagctcagctcagctcaagacgttcaagtg 955
Db 841 CGCTCCGGGGGCTGCGCTGCGACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 900
Qy 956 gaggctcatcatcaaccccccgagtgatgacgacgacgacgacgacgacgacgacgac 1015
Db 901 GAGGCTCATCATCACCCCGAGTGAATCTGTGACAGCCGCGCATCGTGAAAAACCTC 960
Qy 1016 ttaacaaatcagtgatgagcagcagcatttgcggggaattttagaacaacttcaatctc 1075
Db 961 TTAACATCCATGAGCATATTGACGCGCATTTGCGGGATTTTGAACAACTTTCATGTCT 1020
Qy 1076 atggaagcggatcccaagatgaagaagtgtttctcaatccaaatttgaactccaagaac 1135
Db 1021 ATGAGCGGGATGCCAAGTACAAAAGTATTTCTCTCAAAATTTATGATCCAAAGCCA 1080
Qy 1136 agaacaatgacatgctgcgtatgaagctgcagaagccctgcacttcaacgacatagtg 1195
Db 1081 AGAACATGACATTTGCCCTATGAAGCTGCAGAACCTCGATTTCAAGACTATGTA 1140
Qy 1196 aaccagtggtctgcccacccagggcattgtcgcagcagaacagctcgtctgattt 1255
Db 1141 AACCATGTGTCGCCAACCCAGGCAATGATGCTGCAGCCAGAACAGCTCTGCTGATTT 1200
Qy 1256 ccgggtgaggggccacagagaggaaggaagacctcaagaagtgcgaagcgcgcgaag 1315
Db 1201 CCGGCTGGGGGCGCACGAGAGAAAGGAAGACCTCAGAAAGCTGAAAGCTGCCAAGG 1260
Qy 1316 tgcctctcatgagacacagagatgcaacagcagatattgtctatgaaacactgacac 1375
Db 1261 TGCCTCTCATGTGAACACAGATGCAACAGCAGATATGTCTATGACAACTTATCAAC 1320
Qy 1376 cagcagatcgtgtgcggtctcctgcaggggaagcgtcgtattctgcagggtgaaagt 1435
Db 1321 CAGCATGATCTGTGGCGGTCTCTGCAAGGGAACGTCGATTTTGGCAGAGGTGACAGT 1380
Qy 1436 gaaggcctggttcaactcgaaagaacaatatcgtgtgctgataagggatacaagctgg 1495
Db 1381 GAGGCGCTCTGTCTCACTTCAACAACTATCTGTGTGCTGATAGGGGATACAACTGGG 1440
Qy 1496 gttctgctgtgccaagccttaacagacagagagtgatacgggaatgtgatactacgg 1555
Db 1441 GTTCTGCTGTGCAAAAGCTTACAGACCAAGAGTGTACGGGAATGTGATGTTCACGG 1500
Qy 1556 actgatttatacgaataatgaggcagcagctaatccaatcgaatggtctccttcaagt 1615
|||||

```

FT DOMAIN 240 350 CUB.
FT DOMAIN 358 520 MAM.
FT DOMAIN 540 650 CUB.
FT DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 694 787 SRCR.
FT ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 236 BY SIMILARITY.
FT DISULFID 659 671 BY SIMILARITY.
FT DISULFID 666 684 BY SIMILARITY.
FT DISULFID 678 693 BY SIMILARITY.
FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
FT DISULFID 826 842 BY SIMILARITY.
FT DISULFID 926 993 BY SIMILARITY.
FT DISULFID 957 972 BY SIMILARITY.
FT DISULFID 983 1011 BY SIMILARITY.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 170 170 POTENTIAL.
FT CARBOHYD 194 194 POTENTIAL.
FT CARBOHYD 233 233 POTENTIAL.
FT CARBOHYD 263 263 POTENTIAL.
FT CARBOHYD 264 264 POTENTIAL.
FT CARBOHYD 404 404 POTENTIAL.
FT CARBOHYD 456 456 POTENTIAL.
FT CARBOHYD 486 486 POTENTIAL.
FT CARBOHYD 519 519 POTENTIAL.
FT CARBOHYD 550 550 POTENTIAL.
FT CARBOHYD 646 646 POTENTIAL.
FT CARBOHYD 698 698 POTENTIAL.
FT CARBOHYD 722 722 POTENTIAL.
FT CARBOHYD 741 741 POTENTIAL.
FT CARBOHYD 762 762 POTENTIAL.
FT CARBOHYD 864 864 POTENTIAL.
FT CARBOHYD 903 903 POTENTIAL.
FT CARBOHYD 965 965 POTENTIAL.
FT CARBOHYD 166 192 MISSING (IN SHORT ISOFORM).
FT VARSPLIC 808 808 R -> Y (IN REF. 3).
FT CONFLICT 808 808
SQ SEQUENCE 1035 AA; 114887 MW; 5173034A CRC32;

Query Match 24.4% Score 895; DB 1; Length 1035;
Best Local Similarity 36.3% Pred. No. 1.37e-211;
Matches 188; Conservative 84; Mismatches 141; Indels 17; Gaps 16;

```

```

QY 464 CAAKAPRGVGNVAVETDWT 483
ID 11:111111:111111
RESULT 3
ID ENTK_HUMAN STANDARD; PRT: 1019 AA.
AC P96073;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-DUODENUM;
RX MEDLINE: 95234679.
RA KIRAMOTO Y., VEILLE R.A., DONIS-KELLER H., SADLER J.E.;
RT "CDNA sequence and chromosomal localization of human enterokinase,
RT the proteolytic activator of trypsinogen."
RL Biochemistry 34:4562-4568(1995).
[2]
RN SEQUENCE OF 749-1019 FROM N.A.
RP TISSUE-DUODENUM;
RX MEDLINE: 94329561.
RA KIRAMOTO Y., YUAN X., WU Q., MCCOY D.W., SADLER J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains."
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC CARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIVER-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRIPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U09860; AAC50138.1;
CC HSSP: P00763; IDPO.
CC MIM: 226200;
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS01180; CUB; 2.
CC PROSITE: PS00740; MAM_1; 1.
CC PROSITE: PS00740; MAM_2; 1.
CC PROSITE: PS01209; LDLRA_1; 2.
CC PROSITE: PS01209; LDLRA_2; 2.
CC PROSITE: PS00068; LDLRA_2; 2.
CC PFMW: PF00057; 1d1_recept_a; 2.
CC PFMW: PF00089; trypsin; 1.

```


KM Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
 KM Serine protease: Zymogen; Transmembrane; Repeat.
 FT CHAIN 52 117 NON-CATALYTIC M CHAIN (MINI CHAIN).
 FT CHAIN 118 799 NON-CATALYTIC H CHAIN (HEAVY CHAIN).
 FT CHAIN 800 1034 CATALYTIC L CHAIN (LIGHT CHAIN).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 240 349 CUB.
 FT DOMAIN 357 519 MAM.
 FT DOMAIN 359 649 CUB.
 FT DOMAIN 656 694 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 693 786 SRCR.
 FT ACT_SITE 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 986 986 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 986 986 MINISTATE (POTENTIAL).
 FT LIPID 2 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236
 FT DISULFID 658 670
 FT DISULFID 665 683
 FT DISULFID 677 692
 FT DISULFID 787 911
 FT DISULFID 825 841
 FT DISULFID 925 992
 FT DISULFID 956 971
 FT DISULFID 982 1010
 FT DISULFID 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 170 170
 FT CARBOHYD 194 194
 FT CARBOHYD 283 283
 FT CARBOHYD 343 343
 FT CARBOHYD 350 350
 FT CARBOHYD 403 403
 FT CARBOHYD 455 455
 FT CARBOHYD 485 485
 FT CARBOHYD 518 518
 FT CARBOHYD 549 549
 FT CARBOHYD 645 645
 FT CARBOHYD 697 697
 FT CARBOHYD 701 701
 FT CARBOHYD 721 721
 FT CARBOHYD 740 740
 FT CARBOHYD 761 761
 FT CARBOHYD 804 804
 FT CARBOHYD 863 863
 FT CARBOHYD 902 902
 FT CARBOHYD 964 964
 SO SEQUENCE 1034 AA: 114776 MW: 24386471 CRC32:
 Query Match 23.5%; Score 862; DB 1; Length 1034;
 Best Local Similarity 35.0%; Pred. No. 3,79e-202;
 Matches 133; Conservative 89; Mismatches 141; Indels 17; Gaps 16;

OY 344 DIALMLKLOKPLTFNDLVKPYCLPNPBMLOPEOLCWMISGNGATEKERTSEVINAARVLL 403
 DB 951 LSNEKGOOO-MPEXNTENMMACAGEEGIDSCOGDSGFLMCLENNRMVLGAVTSFGYQ 1009
 OY 404 IETORCSRRYIDLLTPAMICAGFLOGNDVSCOGDSGGLVTSKRNIMLIGDTSNGSG 463
 DB 1010 CALPNRPQYARVPEKTEWI 1029
 OY 464 CARKYRQYGVGNVVFETDWI 483
 RESULT 5
 ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-C57BL/6; TISSUE-DUODENUM;
 RA YUAN X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC -1- PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC -1- A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC -1- TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC -1- PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC -1- TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC -1- MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 CC -1- SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to jlicense@isb-sib.ch).
 CC -----
 DR EMBL: U73378; AAB37317.1; -
 DR HSSP: P00763; IDPO.
 DR MGD: MGI:1197523; PRSS7.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS00060; MAM_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS00067; LDLRA_2; 2.
 DR PFAM: PF00057; ldl_recept_a; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00431; CUB; 2.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF00629; MAM; 1.
 DR PFAM: PF01390; SEA; 1.
 DR Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
 KM Serine protease: Zymogen; Transmembrane; Repeat.


```

FT CARBOHYD 127 127
FT CARBOHYD 308 308
FT CARBOHYD 396 396
FT CARBOHYD 453 453
FT CARBOHYD 494 494
FT ACT_SITE 434 434
FT ACT_SITE 483 483
FT ACT_SITE 578 578
FT DISULFID 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 194
FT DISULFID 137 166
FT DISULFID 141 147
FT DISULFID 201 284
FT DISULFID 227 256
FT DISULFID 231 237
FT DISULFID 292 375
FT DISULFID 318 347
FT DISULFID 322 328
FT DISULFID 340 345
FT DISULFID 383 503
FT DISULFID 419 435
FT DISULFID 517 584
FT DISULFID 548 563
FT DISULFID 574 602
SQ SEQUENCE 638 AA; 71369 MW; DAF630DA CRC32;

Query Match 20.7%; Score 762; DB 1; Length 638;
Best Local Similarity 42.9%; Pred. No. 1,15e-173;
Matches 109; Conservative 51; Mismatches 84; Indels 10; Gaps 9;

Db 371 SLRCLNCDNSVCTKSTRIVGCTNSKGEWPMQVSLQVLTNRHLGSGSLGHQVLT 430
Oy 237 SLR-CICGVNLSSROS-RIVGGSALPGAMPQVSLHVO-NV--HYCGSLITPEVIV 291
Db 431 TAACFCGLPLQDVR-I-YSGILNLSDITKDFPSQKEIILHONYKSEGNHDLALIK 488
Oy 292 TAAICVEK-PLNNPMHTAFAGILROSMFYGAGY-OVEKYSHPNDOSKRNIDIALMK 349
Db 489 LQADPLNTEFOKPICLPSKGDSTIYVNCVWTGNGFSEKEKEIIONILOKVIRPLVTNEC 548
Oy 350 LQKPLTFENDLVKPYCLPFGMLQPEOLCMTSGWATEEKEKTSVLAARVLLIETORC 409
Db 549 CKRRO-DKTIORAVACGKKEGKDGCKDGSGPLVCKHNMMPRLVGTSGEGCARREO 607
Oy 410 NSRYVNDLITPAMICAGFLOGVNDSCGDSGGLVTSKNNIMWLLGDTSMGSCAKAYR 469
Db 608 PGVYTKVAEYMDWI 621
Oy 470 PGYGVNVAEYMDWI 483

RESULT 7
ID KAL_RAT STANDARD; PRT: 638 AA.
AC P14272;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
OS PK
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 91129236.
RA BEAUBIEN G., ROSINSKI-CHUPIN I., MATTEI M.-G., MBIKAY M., CHRETIEN M.,
RA SEIDAH N.G.;
RL *Gene structure and chromosomal localization of plasma kallikrein.*;
RL Biochemistry 30:1628-1635(1991).

```

```

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE: 90091743.
RA SEIDAH N.G., LADENHEIM R., MBIKAY M., HAMELIN J., LUTFALLA G.,
RA ROUGEON F., LAZURE C., CHRETIEN M.;
RT "The cDNA structure of rat plasma kallikrein.";
RL DNA 8:563-574(1989).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XII, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63357; AAA74563.1; -
DR EMBL: M63358; AAA74563.1; JOINED.
DR EMBL: M63346; AAA74563.1; JOINED.
DR EMBL: M63347; AAA74563.1; JOINED.
DR EMBL: M63349; AAA74563.1; JOINED.
DR EMBL: M63350; AAA74563.1; JOINED.
DR EMBL: M63351; AAA74563.1; JOINED.
DR EMBL: M63352; AAA74563.1; JOINED.
DR EMBL: M63353; AAA74563.1; JOINED.
DR EMBL: M63354; AAA74563.1; JOINED.
DR EMBL: M63355; AAA74563.1; JOINED.
DR EMBL: M63356; AAA74563.1; JOINED.
DR EMBL: M30282; AAA41463.1; -
DR EMBL: M58590; AAA42069.1; -
DR PIR: A39180; KORTPL.
DR HSP: P00750; IRTF.
DR PROSITE: PS00134; TRYPsin_HIS. 1.
DR PROSITE: PS00135; TRYPsin_SER. 1.
DR PROSITE: PS00495; APPLE. 4.
DR PRAM: PF00024; apple. 4.
DR PRAM: PF00089; trypsin. 1.
DR Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Layer;
KM Duplication; Bradykinin.
FT SIGNAL 1 19
FT CHAIN 20 390
FT CHAIN 391 638
FT REPEAT 20 105
FT REPEAT 110 195
FT REPEAT 200 285
FT REPEAT 291 376
FT DOMAIN 389 621
FT CARBOHYD 127 127
FT CARBOHYD 308 308
FT CARBOHYD 396 396
FT CARBOHYD 453 453
FT CARBOHYD 494 494
FT ACT_SITE 434 434
FT ACT_SITE 483 483
FT ACT_SITE 578 578
FT DISULFID 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 194
FT DISULFID 137 166
FT DISULFID 141 147

```

GENERAL INFORMATION:
APPLICANT: Mortisey, James H.
APPLICANT: Comp. Philip C.
TITLE OF INVENTION: Truncated Tissue Factor and FVIIa or
TITLE OF INVENTION: FVII Activator for Blood Coagulation
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richards, Medlock & Andrews
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: Texas
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04493
FILING DATE: 19930512
CLASSIFICATION:
Prior APPLICATION DATA:
APPLICATION NUMBER: US 07/882202
FILING DATE: 13-MAY-1992
Prior APPLICATION DATA:
APPLICATION NUMBER: US 08/021615
FILING DATE: 19-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Y.
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: ORF B34290CIPC/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1440 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1433
OTHER INFORMATION: /product= "Tissue Factor"
OTHER INFORMATION: /note= "Coding portion of human factor VIII cDNA"
OTHER INFORMATION: /citation= ([1])

[illegible]

Db	840	GGGAGACGACGACCTCAAGCAGCAGCAGCGGGGATGTAGCAGACCGCGCGGGGTGGCCAGATC	899
Qy	1105	attcttcataccaattatgacttccaagacaagaacaatgacattgacgtatgaatg	1164
Db	900	ATATATCCCAAGCAGATACGTATCCCGGGCACCACCAACACAGCATGTGCTCTCCGCTG	959
Qy	1165	cagaaagccttcgacttccaagaactagtgaacaacagatgtctcggccaa-----c	1215
Db	960	CACAGCGCCCTGGTCTCTCACTGACCATGTGTGTCCTCTGCTCTGCCCCGAGACGAGCTTC	1019
Qy	1216	ccaaagcatgtctgtcagccagaacaagctctcgtgattcccggttgggggccaacag	1275
Db	1020	TCTGTAGAGGACGGTGGCCCTTCGTGGCCCTTCTATTTGGTCAGCGGGTGGGCGACCTGCTG	1079
Qy	1276	gagaaagggaaagccctcagaagtgctcgaacgcctcgaaggtctctcatctgaacaacag	1335
Db	1080	GACCGTGGCGCCACGGGCCCTGTGGAGCTTACGTGTCTACACGTGCTCCCGCTATATGACCAG	1139
Qy	1336	agatgc-----aacagcagatatgtctctatgaacaactgatcacaccagcatgact	1386
Db	1140	GACGTGCTGCAGCAGATCACCGAAGGTGGGAGAGTCCCCAAATATACGAGATACATCTTC	1199
Qy	1387	tgtcgcggtctctcgtcaggggaaacgtcgaattcttgcgaaggtgacagtggaagggcctctg	1446
Db	1200	TGTGCGGGTACTTCGGATGTGACACCAAGACACTCTGCAAGGGGGACAGTGGAGGCCACAT	1259
Qy	1447	gtcaacttcgaagaacaatatctcgtgtgctgatatgagggatatacaaaactcgtgggtctcgtcgt	1506
Db	1260	GCCACCCACATACCGGGGACAGTGGTGTACTGTACCGGGGATCGTACGTTGGGGCCAGGGCTGC	1319
Qy	1507	ggcaaaagcttaacagaccaggaagtgtacggaatgtgtagtataccaagcactgattat	1566
Db	1320	GCAACCGTGGGACCACTTTGGGGTGTATACCAAGGGTCTCCAGTACATGAGTGGCTGCA	1379
Qy	1567	cgacaatgagaggcaga	1583
Db	1380	AAGCTCATGGCTCAGA	1396

RESULT 11
US-08-475-845-1
: Sequence 1, Application US/08475845
: Patent No. 5788965
:
: GENERAL INFORMATION:
: APPLICANT: Bernner, Kathleen L.
: APPLICANT: Petersen, Lars C.
: APPLICANT: Hart, Charles E.
: APPLICANT: Hedner, Ulla
: APPLICANT: Brengendand, Claus
: TITLE OF INVENTION: Modified Factor VII
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crews
: STREET: One Market Plaza, Steuart Street Tower
: CITY: San Francisco
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 94105-1492
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.24
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,845
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/327,690
: FILING DATE: 24-OCT-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
:

PRIOR APPLICATION DATA: 08/327,690
APPLICATION NUMBER:
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

RESULT 9
ID FAIL_HUMAN STANDARD; PRT; 625 AA.
AC P03951;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN ANTECEDENT) (PRA).
GN F11.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86243360.
RA FUJIKAWA K., CHUNG D.W., HENDRICKSON L.E., DAVIE E.W.;
RT "Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein.";
RT Biochemistry 25:2417-2424(1986).
RL [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 88107663.
RA ASAKAI R., DAVIE E.W., CHUNG D.W.;
RT "Organization of the gene for human factor XI.";
RT Biochemistry 26:7221-7228(1987).
RL [3]
RN PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE; 91152017.
RA MCNULLEN B.A., FUJIKAWA K., DAVIE E.W.;
RT "Location of the disulfide bonds in human coagulation factor XI: the presence of tandem apple domains.";
RT Biochemistry 30:2056-2060(1991).
RL [4]
RN VARIANT LEU-301.
RP MEDLINE; 90046656.
RX ASAKAI R., CHUNG D.W., RATNOFF O.D., DAVIE E.W.;
RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi Jews is a bleeding disorder that can result from three types of point mutations.";
RT Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).
RL [5]
RN VARIANT LEU-301.
RP MEDLINE; 92190478.
RX MEIJERS J.C., DAVIE E.W., CHUNG D.W.;
RT "Expression of human blood coagulation factor XI: characterization of the defect in factor XI type III deficiency.";
RT Blood 79:1435-1440(1992).
RL -1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-1-ALA AND ARG-1-VAL BONDS IN FACTOR IX TO FORM FACTOR IXA.
CC -1- SUBUNIT: HOMODIMER, LINKED BY A DISULFIDE BOND. AFTER ACTIVATION THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.
CC -1- PTM: ACTIVATED BY FACTOR XIIa (OR XII), WHICH CLEAVES EACH POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH MOLECULAR WEIGHT (HMW) KININOGEN.
CC -1- DISEASE: DEFECTS IN F11 ARE A CAUSE OF A BLOOD COAGULATION ABNORMALITY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN ASHKENAZI JEWS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M13142; AAAS2487.1; -
DR EMBL; M20218; AAAS1985.1; -
DR EMBL; M18296; AAAS1985.1; JOINED.
DR EMBL; M2184; AAAS1985.1; JOINED.
DR EMBL; M18298; AAAS1985.1; JOINED.
DR EMBL; M18300; AAAS1985.1; JOINED.
DR EMBL; M18301; AAAS1985.1; JOINED.
DR EMBL; M18302; AAAS1985.1; JOINED.
DR EMBL; M18303; AAAS1985.1; JOINED.
DR EMBL; M18304; AAAS1985.1; JOINED.
DR EMBL; M19417; AAAS1985.1; JOINED.
DR EMBL; M20217; AAAS1985.1; JOINED.
DR PIR; A27431; KFHU1.
DR HSSP; P00763; IDPO.
DR MIM; 134540; -
DR MIM; 264900; -
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PFAM; PF00024; apple; 4.
DR PFAM; PF00089; trypsin; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Blood coagulation;
KW Duplication; Signal; Disease mutation.
FT SIGNAL 1 18
FT CHAIN 19 387 COAGULATION FACTOR XIA HEAVY CHAIN.
FT CHAIN 388 625 COAGULATION FACTOR XIA LIGHT CHAIN.
FT REPEAT 19 104 APPLE 1.
FT REPEAT 109 194 APPLE 2.
FT REPEAT 199 284 APPLE 3.
FT REPEAT 290 375 APPLE 4.
FT DOMAIN 384 625 CATALYTIC.
FT CARBOHYD 90 90
FT CARBOHYD 126 126
FT CARBOHYD 353 353
FT CARBOHYD 450 450
FT CARBOHYD 491 491
FT ACT_SITE 431 431
FT ACT_SITE 480 480
FT ACT_SITE 575 575
FT ACT_SITE 20 103
FT DISULFID 29 29 WITH A CYS RESIDUE.
FT DISULFID 46 76
FT DISULFID 50 56
FT DISULFID 110 193
FT DISULFID 136 165
FT DISULFID 140 146
FT DISULFID 200 283
FT DISULFID 226 255
FT DISULFID 230 236
FT DISULFID 291 374
FT DISULFID 317 346
FT DISULFID 321 327
FT DISULFID 339 339
FT DISULFID 380 500
FT DISULFID 416 432 INTERCHAIN.
FT DISULFID 514 581 INTERCHAIN (BETWEEN HEAVY AND LIGHT CHAINS).
FT DISULFID 545 560
FT DISULFID 571 599
FT VARIANT 301 301
FT CONFLICT 226 226 F -> L (IN F11 DEFICIENCY).
FT SEQUENCE 625 AA; C -> S (IN REF. 2).
SQ 70109 MW; C46A5F74 CRC32;
Query Match 18.9%; Score 694; DB 1; Length 625;
Best Local Similarity 42.6%; Pred. NO. 1,966-154;
Matches 100; Conservative 43; Mismatches 86; Indels 6; Gaps 5;
Db 386 PRIVGASVNGEMPMOYTLTTSPTORHLCGSGTICQWILTAHCYF-GVESPKILRV 444
:|||||:::|||||:|:|||||:|||||:|:|
253 SRVGGESALPGAMPQVSLHVON-V--HYCGSGITPEWIVTAHCVEKPLNPMHMTA 309

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF030065; AAB84221.1; -
 DR HSP; P00763; IDPO.
 DR GSD; MG11196620; HPN.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 DR PFAM; PF00089; trypsin; 1.
 KM Hydroxylase; Serine protease; Transmembrane; signal-anchor.
 FT CHAIN 1 161
 FT CHAIN 1 16
 FT DOMAIN 1 16
 FT TRANSMEM 17 43
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT CATALYTIC.
 FT ACT_SITE 162 416
 FT ACT_SITE 202 416
 FT ACT_SITE 256 416
 FT ACT_SITE 352 416
 FT ACT_SITE 352 416
 FT DISULFID 152 276
 FT DISULFID 187 203
 FT DISULFID 321 337
 FT DISULFID 348 380
 FT DISULFID 111 111
 FT CARBOHYD 111 111
 FT SEQUENCE 416 AA; 44739 MW; C1297CCE CRC32;

Query Match 18.0%; Score 660; DB 1; Length 416;
 Best Local Similarity 33.7%; Pred. No. 7,22e-145; Indels 20; Gaps 15;
 Matches 121; Conservative 76; Mismatches 142;

DB 53 VQSPGDSRLAVDKTEGTWRLCCSSRNARVAGLCEGEMGLRALHSELIVRTAGANG 112
 113 TSQFPCVDEGLPLAQLLDVIVSCDPRGRFLATQCDDGR - KLPVD - RIVGQSSSL 170
 149 VRLYGFNFILQVSSQKSNHPVCDDMMENYGRACRDWGRKNNFYSSQGIYDDSGSTS 208
 209 FMKLN-TSAGNVDLYKKY-HS-DACS-SKA-VYSLRIACVNNSSQSGIVG-ESAL 262
 171 -GAMPQVSLRYDGTSLCGSLGSDVNLAAHGFPERNRVLSMRVAFAGAVARTSP-HA 228
 263 FGAMPQVSLHVNVAHVGSGSIITPEWIVPAHCEKPLNNPMHTAFAGILRGSFMYG 322
 229 VQGVQAVIYHGYLEPRDPTIDENSNDIALVHLSSLPLEYIQVCLPAAGQALVDGK 288
 323 AGVQEVKIVHPN---DSKTK-N-NDIALMKLQKPLTFNDLVKPCPLRPGMMQLEQ 376
 289 VCVVTGKNTQFTGQAMVLOEARVPIISNEVCNSPFTYQIQPKKFCACIPGSGIDAC 348
 377 LCIWISGATLEEGKTSSEVNAALVLIETQRCNSRYVDLITPAMICAGFLQGNVDS 436
 349 QGSGGFEVEDSISGSRMRIGIVSMGTGALARKRGVYTKTDRFEMFKIKTHS 407
 437 QGSGGGLV--TSKNNT--WMLIGDTSMGSCAKAYRPGVGNWFTDWTIRMRADG 491

RESULT 12
 ID EPIIT_MOUSE STANDARD; PRT; 902 AA.
 AC P56677;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE EPIITIN (EC 3.4.21.-).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C.B.17SCID; TISSUE-THYMUS;

RA KIM M.G., CHEN C., LYO M.S., CHO E., PARK D., KOZAK C., SCHWARTZ R.H.;
 RT Cloning and chromosomal mapping of a gene encoding a novel mouse type
 RT II membrane serine protease, epithelin, containing four LDL receptor
 RT modules and two CUB domains."
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF042822; AAD02230.1; -
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KM Signal-anchor; Glycoprotein; Hydroxylase; Serine protease;
 FT TRANSMEM 1 55
 FT DOMAIN 56 76
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT CUB 1.
 FT CUB 2.
 FT LDL-RECEPTOR CLASS A 1.
 FT LDL-RECEPTOR CLASS A 2.
 FT LDL-RECEPTOR CLASS A 3.
 FT LDL-RECEPTOR CLASS A 4.
 FT CATALYTIC.
 FT ACT_SITE 656 656
 FT ACT_SITE 711 711
 FT ACT_SITE 711 711
 FT ACT_SITE 805 805
 FT CARBOHYD 107 107
 FT CARBOHYD 302 302
 FT CARBOHYD 365 365
 FT CARBOHYD 421 421
 FT CARBOHYD 489 489
 FT CARBOHYD 772 772
 FT SEQUENCE 902 AA; 99645 MW; 422D3BBF CRC32;

Query Match 17.9%; Score 658; DB 1; Length 902;
 Best Local Similarity 40.4%; Pred. No. 2,63e-144; Indels 11; Gaps 10;
 Matches 97; Conservative 57; Mismatches 75;

DB 602 CDCGLSFTKQARVVGNTNDEGEMPOVSLHALGQHLGASLSIPDVLVAHCFOD 661
 243 CGVNLN-SRQSRIVGESALPGAMPQVSLH-VQNVHVGSGSIITPEWIVTAHCV--E 288
 662 KNEFYSDYTMATLGLLDOSKRSASGOYELKLRITTHSFENFTDYDIALLEKSV 721
 299 KPLNMP-WH-WTAFAGILRGSFME-YGA-GYOYEVKIVSHPNYDSTKNNIDIALMKLQKPL 354
 722 EYTVVVRPICLPDATHVFPAGKAIWVGHTKGGTGAALQAGEIRVINTQTCED-LM 780
 355 TFNDLVKPVCLPFGMMQLEPQOLCHISGWGTEKGTSEVLNAKALLIETQRCNSRYV 414
 781 -PQGITPRMKVGLSGGVSCQDGGPLSSABKDRMFGAGVYSWGECAQNRNKGIV 839
 415 YDNLITPAMICAGFLQGNVDSQCGDGGPLVTSKNNTIW-WLIGDTSMGSCAKAYRPGV 473

RESULT 13
 ID HEP5_RAT STANDARD; PRT; 416 AA.
 AC Q05511;

```

APPLICANT: Hart, Charles E.
APPLICANT: Petersen, Lars C.
APPLICANT: Hedner, Ulla
APPLICANT: Rasmussen, Mirella E.
TITLE OF INVENTION: Modified Factor VII
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,003
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 90-0707
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 28..1420
OTHER INFORMATION: /codon_start= 28
OTHER INFORMATION: /product= "Factor VII"
US-08-871-003-1

```

```

Query Match          5.58; Score 95.4; DB 4; Length 2422;
Best Local Similarity 47.98; Pred. No. 1.2e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

QY 865 cgcacagagcaggaatgtggtgagcagagcgcgtcccggtggccttgccctgcaagtc 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 599 CCCCAAGCGCAATTGGGGGGGCAAGGTGTGCCCAAGGGAGTGTCCATGCGAGGTC 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 925 agcctgcaagtcagaaagtcacgctgtgagaggtccatcatcaaccgccagatgac 984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 659 CTGTGTTGGTGAATGAGCTCACTGTGTGTGGGGGACCCCTGATCAACACCATCTGGGTG 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 985 gtgacagcgcgcgcctcggtgagaaaccccttaacaatcattgcatgtgacggcatt 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 719 GTCTCCGGCGCCCACTGTTTCGACAAATCAAGAACTGGAGAACTGTATCGCGTGTGCTG 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1045 gcggggattttgagacaactcttcattgttcattgagccgagataccaagtgaagaa 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 GCGGAGCAGCAGCTCAGCAGCAGCAGCGGGGATGAGCAGAGCGCGGTGGCGCAGGTC 838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1105 attctcatccaatttgactccaagaacaagaacattgacattgagtgagagtcg 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 ATCATCCCAAGCAGTACGTCCCGGGACCAACCAACGACATCGCGCTCTCCGCTG 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1165 cagaaagcctcgtacttcaagagcattgtgaaccagtggtgtcggccaa-----c 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 899 CACAGCCCGTGGTCTCTCACTGAACATGTGTGGCCCTGTGCTGCCCGCAAGGACGTTT 958
QY 1216 ccaggcatgatctgtcagccagaacagctctgtgatttcgggtgtggggccaccag 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 959 TCTTAGAGAGAGGCTGCGCTTCTGCGCTTCTCATTTGCTCAGCGGGGCGCAGTGTG 1018
QY 1276 gagaagaaggaaacctcagaagtgtgtgaagctgcgcaaggtgtcttcattgagacag 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1019 GACCGTGGCCGACAGGCGCTGGAGCTCATGTGCTCAACGTGCCCGCGGTGATGCCAG 1078
QY 1336 agatgc-----aacagcagatatgtctatgacacactgtatcaaccagccatgac 1386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1079 GACTGCTGCAAGCAGTACCGGAGGTGGAGACTCCCAAAATTCACGAGTACATGTTT 1138
QY 1387 tgtgcggtctcctgcaggggaacgttcgattctgcaggggtgacagtgaggccctcg 1446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1139 TGTGCGGCTACTCTCGATGACACCAAGACTCTCGCAAGGGGAGACGTGAGGCGCCACAT 1198
QY 1447 gtcacttcgaaagaacaaatattctgtgtgtatagggagatacaagctgggttctgtc 1506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1199 GCCACCCACTACCGGGGCGACGTGTACTGACGGGCACTGCTCAGCTGGGCGCAGGCTGC 1258
QY 1507 gccaaagcttacagaccagagtgtaaggaatgtgattgtatcaacgagactgtattat 1566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1259 GCAACCGTGGGCGCACTTGGGGGTGTATACCAAGGCTCTCCCAAGTACATGAGTGCCAA 1318
QY 1567 cgaacaattgaggcaga 1583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1319 AAGCTCATGCGCTCAGA 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: April 22, 2000, 06:01:05
 Job time: 2009 sec

DR PFAM: PF00051: kringle; 5.
 DR PFAM: PF00089: trypsin; 1.
 DR Hydrolase: Serine protease; Plasma: Glycoprotein; Fibrinolysis;
 KW Tissue remodeling; Blood coagulation; Kringle; Zymogen; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 812 PLASMINOGEN.
 FT CHAIN 20 581 PLASMINOGEN HEAVY CHAIN A.
 FT PEPTIDE 20 97 ACTIVATION PEPTIDE.
 FT CHAIN 98 581 PLASMINOGEN SHORT FORM OF CHAIN A.
 FT CHAIN 98 7436 ANGIOSTATIN.
 FT CHAIN 582 812 PLASMINOGEN LIGHT CHAIN B.
 FT DOMAIN 103 181 KRINGLE 1.
 FT DOMAIN 184 262 KRINGLE 2.
 FT DOMAIN 275 352 KRINGLE 3.
 FT DOMAIN 377 454 KRINGLE 4.
 FT DOMAIN 481 560 KRINGLE 5.
 FT DOMAIN 582 812 CATALYTIC.
 FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 667 667 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 762 762 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 49 73 BY SIMILARITY.
 FT DISULFID 53 61 BY SIMILARITY.
 FT DISULFID 103 181 BY SIMILARITY.
 FT DISULFID 124 164 BY SIMILARITY.
 FT DISULFID 152 176 BY SIMILARITY.
 FT DISULFID 185 262 BY SIMILARITY.
 FT DISULFID 188 316 BY SIMILARITY.
 FT DISULFID 206 245 BY SIMILARITY.
 FT DISULFID 224 257 BY SIMILARITY.
 FT DISULFID 234 352 BY SIMILARITY.
 FT DISULFID 296 335 BY SIMILARITY.
 FT DISULFID 324 347 BY SIMILARITY.
 FT DISULFID 377 454 BY SIMILARITY.
 FT DISULFID 398 437 BY SIMILARITY.
 FT DISULFID 426 449 BY SIMILARITY.
 FT DISULFID 481 560 BY SIMILARITY.
 FT DISULFID 502 543 BY SIMILARITY.
 FT DISULFID 531 555 BY SIMILARITY.
 FT DISULFID 568 687 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 578 586 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 609 625 BY SIMILARITY.
 FT DISULFID 701 768 BY SIMILARITY.
 FT DISULFID 731 747 BY SIMILARITY.
 FT DISULFID 758 786 BY SIMILARITY.
 FT SEQUENCE 812 AA: 90846 MW: 6C120F7A CRC32:
 Query Match 17.3%; Score 634; DB 1; Length 812;
 Best Local Similarity 41.4%; Pred. No. 1,41e-137;
 Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;

AC P80010;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PLASMINOGEN (EC 3.4.21.7) (FRAGMENT).
 GN PLG.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Perissodactyla; Equidae; Equus.
 RP [1]
 RP SEQUENCE.
 RC TISSUE-PLASMA;
 RX MEDLINE; 92052077.
 RA SCHALLER J., STRAUB C., KAEMPFER U., RICKLI E.E.;
 RT "Complete amino acid sequence of equine miniplasminogen";
 RL Protein Seq. Data Anal. 4:69-74(1991).
 CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
 CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
 CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
 CC AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
 CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
 CC ACTIVATOR. COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
 CC AS C1 AND C5, IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
 CC LAMININ AND VON WILLEBRAND FACTOR.
 CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
 CC ACTIVATORS. BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
 CC FIBRIN. ACTIVATED WITH CATALYTIC AMOUNTS OF STREPTOKINASE.
 CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
 CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS AT LEAST 1 KRINGLE REGIONS.
 CC PIR: S17527; S17527.
 DR HSSP: P00747; SHRG.
 DR PROSITE: P500021; KRINGLE_1; 1.
 DR PROSITE: P500134; TRYPSIN_HIS; 1.
 DR PROSITE: P500135; TRYPSIN_SER; 1.
 DR PROSITE: P50070; KRINGLE_2; 1.
 DR PFAM: PF00051; kringle; 1.
 DR PFAM: PF00089; trypsin; 1.
 KW Hydrolase; Serine protease; Plasma: Glycoprotein; Fibrinolysis;
 KW Tissue remodeling; Blood coagulation; Kringle; Zymogen.
 FT NON_TER 1 1
 FT CHAIN 109 338 HEAVY CHAIN A.
 FT DOMAIN 9 88 LIGHT CHAIN B.
 FT DISULFID 109 338 KRINGLE 5.
 FT DISULFID 9 88 CATALYTIC.
 FT DISULFID 30 71 BY SIMILARITY.
 FT DISULFID 59 83 BY SIMILARITY.
 FT DISULFID 95 213 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 105 113 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 135 151 BY SIMILARITY.
 FT DISULFID 227 294 BY SIMILARITY.
 FT DISULFID 257 273 BY SIMILARITY.
 FT DISULFID 264 312 BY SIMILARITY.
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 288 288 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 157 157 STREPTOKINASE-BINDING SITE (PROBABLE).
 FT SITE 191 191 STREPTOKINASE-BINDING SITE (PROBABLE).
 FT SITE 269 269 STREPTOKINASE-BINDING SITE (PROBABLE).
 FT SITE 282 282 SITE OF SUBSTRATE SPECIFICITY (BY SIMILARITY).
 FT SEQUENCE 338 AA: 37132 MW: 81DD0C34 CRC32:
 Query Match 16.8%; Score 617; DB 1; Length 338;
 Best Local Similarity 38.2%; Pred. No. 8,01e-133;
 Matches 97; Conservative 66; Mismatches 79; Indels 12; Gaps 12;

 W P S R E H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msearch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 23:28:21 2000; Maspar time 27.50 Seconds
 842.259 Million cell updates/sec

Tabular output not generated.

Title: >US-09-323-597-2
 Description: (1-491) from US09323597.pep
 Sequence: 3673
 1 MALNGSPPAICPYENHGY.....VYGNVVFETDWIRQMRADG 491

Scoring table: PAM 150
 Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: plr62
 1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 47.373; Variance 79.961; scale 0.592

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Pred. No.
1	895	24.4	1035 1	AA3090 enteropeptidase (EC 3	3.16e-180
2	890	24.2	1019 1	A56318 enteropeptidase (EC 3	5.25e-179
3	862	23.5	1034 1	A53663 enteropeptidase (EC 3	3.55e-172
4	762	20.7	638 1	KOHUP plasma kallikrein (EC	6.82e-148
5	740	20.1	638 1	KORTPL plasma kallikrein (EC	1.42e-142
6	733	20.0	1113 2	JF0315 low-density lipoprote	6.93e-141
7	720	19.6	638 1	KOMSPPL plasma kallikrein (EC	9.45e-138
8	694	18.9	625 1	KRHUI coagulation factor XI	1.71e-131
9	680	18.5	417 1	S00845 hepsin (EC 3.4.21.-)	3.96e-128
10	652	17.8	416 1	S33777 hepsin (EC 3.4.21.-)	2.04e-121
11	634	17.3	812 1	PLMS plasmin (EC 3.4.21.7)	4.10e-117
12	617	16.8	812 1	PLMS plasmin (EC 3.4.21.7)	4.65e-113
13	613	16.7	790 1	PLPG plasmin (EC 3.4.21.7)	4.18e-112
14	608	16.6	810 2	B50848 plasmin (EC 3.4.21.7)	6.48e-111
15	597	16.3	810 1	PLHU mast cell tryptase pr	1.12e-110
16	597	16.3	810 1	PLHU plasmin (EC 3.4.21.7)	4.63e-108
17	583	15.9	267 2	B61545 plasmin (EC 3.4.21.4)	5.65e-105
18	582	15.8	273 2	A47246 tryptase (EC 3.4.21.5)	9.75e-105
19	582	15.8	273 2	A47246 tryptase (EC 3.4.21.5)	1.49e-103
20	577	15.7	276 2	A38654 mast cell proteinase	2.29e-102
21	572	15.6	275 2	C35863 tryptase (EC 3.4.21.5)	2.29e-102
22	572	15.6	275 2	C35863 tryptase (EC 3.4.21.5)	2.29e-102
23	572	15.6	275 2	A35863 tryptase (EC 3.4.21.5)	2.29e-102

24	568	15.5	421 1	S11674	acrosin (EC 3.4.21.10)	2.02e-101
25	570	15.5	810 2	I46260	plasmin (EC 3.4.21.7)	6.80e-102
26	569	15.5	812 1	PLBO	plasmin (EC 3.4.21.7)	1.17e-101
27	564	15.4	263 2	A21195	chymotrypsin (EC 3.4.	1.79e-100
28	564	15.4	343 1	A57014	prolactin (EC 3.4.21.	1.79e-100
29	561	15.3	274 2	JC4171	trypsin (EC 3.4.21.5)	9.16e-100
30	558	15.2	245 1	KYBOA	chymotrypsin (EC 3.4.	4.69e-99
31	557	15.2	275 2	S40007	trypsin (EC 3.4.21.4)	8.08e-99
32	555	15.1	245 1	KYBOB	chymotrypsin (EC 3.4.	2.40e-98
33	554	15.1	263 2	A31299	chymotrypsin (EC 3.4.	4.13e-98
34	546	14.9	264 2	I38136	chymotrypsin-like pro	3.19e-96
35	548	14.9	274 2	S35339	trypsin (EC 3.4.21.4)	1.08e-96
36	548	14.9	275 2	S40005	trypsin (EC 3.4.21.4)	1.85e-96
37	547	14.9	415 1	A34170	acrosin (EC 3.4.21.10)	3.19e-96
38	546	14.9	655 1	A46688	hepatocyte growth fac	9.45e-96
39	544	14.8	436 2	JX0172	acrosin (EC 3.4.21.10)	8.28e-95
40	540	14.7	277 2	S35340	trypsin (EC 3.4.21.4)	2.45e-94
41	538	14.6	263 1	KYRTB	chymotrypsin (EC 3.4.	2.45e-94
42	538	14.6	263 2	S47537	chymotrypsin (EC 3.4.	1.87e-92
43	530	14.4	456 1	KXBO	protein C (activated)	3.22e-92
44	529	14.4	786 1	A47547	serine proteinase st	1.87e-92
45	530	14.4	1420 2	A32869	apolipoprotein(a) (EC	1.87e-92

ALIGNMENTS

RESULT 1
 ENTRY A43090 #type complete
 TITLE enteropeptidase (EC 3.4.21.9) precursor - bovine
 ALTERNATE_NAMES enterokinase
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS
 REFERENCE A43090
 #authors Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:7588-7592
 #title Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains.

#cross-references MUID:94329561
 #accession A43090
 #status nucleic acid sequence not shown; translated from GB/EMBL/DBJ

REFERENCE
 #molecule_type mRNA
 #residues 1-1035 #label KIT
 #cross-references GB:U09859; NID:9746410; PIDN:AA40026.1; PID:9746411
 #experimental_source small intestine
 #accession A48874

#authors Lavallo, E.R.; Rehmulla, A.; Racle, L.A.; DiBalsio, E.A.; Ferenc, C.; Grant, K.L.; Light, A.; McCoy, J.M.
 #journal J. Biol. Chem. (1993) 268:23311-23317
 #title Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase.
 #cross-references MUID:94043122

#molecule_type mRNA
 #residues 801-1035 #label LAV
 #cross-references GB:U19663; NID:9416131; PIDN:AA16035.1; PID:9416132
 #note parts of this sequence, including the amino end of the mature protein, were confirmed by protein sequencing
 #accession A61436

REFERENCE A61436
 #authors Light, A.; Janska, H.
 #journal J. Protein Chem. (1991) 10:475-480
 #title The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
 #cross-references MUID:92189715
 #accession A61436
 #molecule_type protein
 #residues 801-807, 'Y', 809-827 #label LIG
 #comment The mechanism of association with the membrane of the intestinal brush border is unknown. The sequence is compatible with amino-terminal myristoylation of the heavy chain or with type II


```

DB 694 TGTAGCCGACCTCAACCCCATGCTGTGCAATCTGGGGTTCAAGCTGTGATCTATCATCG 753
OY 1098 aaagctattctcatcacaattatgactcacaagacacaatgacatgctgcgtatc 1157
DB 754 GGGCTACTCTTCCCTTTCGAGACCTTACTATCGAAGAAAGCAAGATGACTTTCCTTGGT 813
OY 1158 gaagctgcagaagcctctgacttcaacagacactgtaagaaacagctgtctgcacaccc 1217
DB 814 CCACCTCTCTAGCTCCCTGCTCTCAACAGAAATCATTCACACCCAGTGTGTCTCCCTGCTGC 873
OY 1218 aggcatagactgctcagcagcagaacagctctgctgattccgggtggggggccacacagga 1277
DB 874 GGGACAGGCGCTCTGGAGTGGCAAGGCTGTACTGTGATCCGCTGGGGTAACACACAGTT 933
OY 1278 gaagaggaagactcagaagctgctgaaagctgcgaagctctctcattgtagacacagag 1337
DB 934 CTATTGGCCACACAGGCTATGTGTGCTCCAGAGGCCCGGCTTCCCATCTATAGCAACAGAT 993
OY 1338 atgcacacagcagatatagtctatgacaaactgatacacaacacagcatgactctgcggctc 1397
DB 994 TTGCACACAGCCCGGACTTCTTACGGGAATCAGATCAAGCCCAAGATGTTCTGTCTGCTGA 1053
OY 1398 cctgcaggggaaagctgactctctgcaaggtgtgacagtgagtgagggcctctgtt----- 1449
DB 1054 TCTGTAGGGGTGATTTGATGCTGTCAGGCGCACAGTGGAGGCCCTTGTGTGTGTAAGA 1113
OY 1449 ----cacttcgaagacacatactctgctgctgataaggaatacaagctgggtctgctgctg 1505
DB 1114 CACATCTCTGGGACATCAAGGTGGCGGCTATGTGGCATTTGAAGCTGGGGTACGGGCTG 1173
OY 1506 tgcacaaagcttacaagacagagagtgtaacgggaatgtatgattcaacggactgattat 1565
DB 1174 TGTCTTGGCCCGGAGAGCAGAGATGTACACCAAGTCACTGACTCCGGAGATGATCTT 1233
OY 1566 tcgaacaatgagggcgagacgctaatcca 1594
DB 1234 CAAGGCCATTAAGACTCTACTCCGAAAGCA 1262

RESULT 7
VS9136
ID VS9136 standard; DNA: 980 BP.
AC VS9136;
DE 07-JAN-1999 (first entry)
DE Nucleotide sequence of SP0031A, a homologue of HELA2.
KW Serine protease; regulation; cell activity; viability; HELA2; ATC2;
KW BCOM; testis; fertility; suppressor; testicular germ cell cancer;
KW seminoma; testis-specific expression; antitumour; sperm development;
KW infertility; human; chromosome 16p13.3; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 3..899 /tag a
FT /product= SP0031A
FT /note="sequence contains one internal stop codon"

MO9836054-A1.
PD 20-AUG-1998.
PD 13-FEB-1998: AU0085.
PR 18-NOV-1997: AU-000422.
PR 13-FEB-1997: AU-005101.
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
PI Antalis TM, Hooper JD;
DR MPI: 98-480768/41.
DR P-PSDB: W77304.
PT New serine protease(s) and kinase involved in regulating cell
PT activity and viability - particularly the testis-specific protease
PT HELA2 used for modulation of fertility and as tumour suppressor
PS Example 15; Fig 20C; 167pp; English.
CC VS9134-36 represent HELA2 homologues. The genes are found in a cluster
CC on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
CC homology to serine proteases. The protein is involved in or associated
CC with regulation of cell activity and/or viability. Administration of
CC recombinant HELA2 (also called testisin) is used to increase fertility.

```

```

CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
CC testicular germ cell cancers (seminoma) and is also expressed in some
CC non-testicular cancers (of colon, pancreas, prostate and ovary). So is
CC a marker/potential therapeutic target for cancer. The promoter from the
CC HELA2 gene is useful for testis-specific expression of other genes,
CC e.g. for gene therapy or modulation of fertility. Drugs that block
CC activity of HELA2 should have antitumour activity (other than in
CC testis) while in testis recombinant HELA2 should stop growth of tumours
CC and normalise sperm development (eliminating the need for orchiectomy).
CC Identification of mutant forms of HELA2 can be used to diagnose
CC infertility.
SQ Sequence 980 BP: 192 A; 325 C; 289 G; 174 T;

Query Match 6.9%; Score 120; DB 1; Length 980;
Best Local Similarity 50.3%; Pred. No. 3..8e-24;
Matches 371; Conservative 0; Mismatches 340; Indels 27; Gaps 2;

OY 858 cttaagccgcagagacagatattgtggcgagagcgccctccggggcctgcccgcg 917
DB 11 CCCACAGATGCTCAACCGAATGCTGGCGGCAGACAGCAGAGGCGGAGTGGCCCTG 70
OY 918 gcaagctcagctcgcagctcagagaaagctccagctgctcggaggctcattcaccacccga 977
DB 71 GCAAGTCAAGCATCCACGCGCACCGAAGCCACTTCTCGGGGGCAGCTCATCGCGAGCA 130
OY 978 gtgatactgtacagccgcacactgcgtggaaaaaccttcaacaatcctatgcatgagac 1037
DB 131 GTGGGCTCTGACGGCGGACGCTGCTCCGCAACACTCTGTGAACGCTCCCTGTACCAAGG 190
OY 1038 ggcatttcggggattttgagacatcttcaatgcttcatgtctatgagacggataccagataga 1097
DB 191 CCTGTGGGGGCAAGCAGCTACTGTGACCGCGGACCAACACGCTATGTATGCCCGGGTGA 250
OY 1098 aaagctattctcatcacaattatgactcacaagacacaatgacatgctgcgtatc 1157
DB 251 GCAAGGTGAGAGCAACCCCTGTACCAAGGACAGGCTCTCAGCGCTGACGTGACCTGTGT 310
OY 1158 gaagctgcagaagcctctgacttcaacagacactgtaagaaacagctgtctgcacaccc 1217
DB 311 GGAAGCTGAGAGCAACGATGCGCTTACACAAATTAATCTCTCCCGTGTGCTGCTGACCC 370
OY 1218 aggcatagactgctcagcagcagaacagctcctgattccgggtggggggccacagga 1277
DB 371 CTGGGTGATCTTGAACAGCGGACATGAATCTCTGCTGCTGCTGCGGAGAGCCCAAGTA 430
OY 1278 gaagaggaagactcagaagtgctgtaagctgcacaa-----gtgctctcattgaagc 1331
DB 431 GGAAGACTCTCTCCCGAACCCGGGATCTCTGCAGAAACTCGCTGTCCCATCTATCGACAC 490
OY 1332 acagagatgcacagcagatattgctatgacaactct-----at 1370
DB 491 ACCCAAGTGCACCTGCTCTACAGCAAGACACACGAGTTTGGCTACCAACCAAAACAT 550
OY 1371 cacaacacagatattctgctcggtctctcagggggaacgtgattcttgcagaggtga 1430
DB 551 CAAGAAATGACATGCTGTGCGCCCGCTTCGAGAGGCGCAAGAGAGCTCTCAAGCGCA 610
OY 1431 cagtgagagggccctctgctcctcgaagaacaaatattctgtgctgtagtggaatacaag 1490
DB 611 CTGGGGGCGCCCTGCTGTGCTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 670
OY 1491 ctgggtctcgtctgctgcacaagcttacaagacagagagtgtaacgggaatgtgattat 1550
DB 671 CTGGGGTGAAGGCTGTGCGCGCGCAAGACCGCCAGAGTGTATATCGGTACCGCCCA 730
OY 1551 cagcagtgattatcgcg 1568
DB 731 CCACAACCTGGATCATCG 748

RESULT 8
O70104

```

```

#molecule_type mRNA
##residues 1-638 ##label SEI
##cross-references GB:M30287; NID:g205010; PID:AAA4163.1; PID:g205011
part of this sequence, including the amino ends of both
the heavy and light chains, was confirmed by protein
sequencing

REFERENCE
#authors S06851
#journal Blochm. Biophys. Acta (1989) 999:103-110
#title Rat plasma kallikrein: purification, NH(2)-terminal
#cross-references MURD:90089457
#accession S06851
##molecule_type protein
##residues 20-45;391-413 ##label PAO
#authors Seldah, N.G.; Ladenheim, R.; Mlikay, M.; Hamelin, J.;
#journal Lutfalla, G.; Rougeon, R.; Lazure, C.; Chretien, M.
#title DNA Cell Biol. (1988) 8:563-574
#accession I53041
#status translated from:GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-638 ##label RES
##cross-references GB:M58590; NID:g206721; PID:AAA42069.1; PID:g206722
COMMENT This protein, synthesized in the liver, circulates as a noncovalent
complex with high molecular weight (HMW) kininogen.
COMMENT The zymogen is activated by factor XIIa, which cleaves the molecule
into a light chain, which contains the active site, and a heavy
chain, which associates with HMW kininogen. These chains are
linked by one or more disulfide bonds.

GENETICS
#gene PK
CLASSIFICATION #superfamily coagulation factor XI; trypsin homology
KEYWORDS blood coagulation; duplication; fibrinolysis; glycoprotein;
hydrolyase; inflammation; liver; plasma; serine proteinase;
zymogen

FEATURE
1-19 #domain signal sequence #status predicted #label SIG\
20-390 #product plasma kallikrein heavy chain #status
experimental #label MAT1\
20-109 #domain apple repeat #label AP1\
110-199 #domain apple repeat #label AP2\
200-289 #domain apple repeat #label AP3\
291-380 #domain apple repeat #label AP4\
391-638 #product plasma kallikrein light chain #status
experimental #label MAT2\
391-621 #domain trypsin homology #label TRY\
21-104;47-77,51-57,
111-194;137-166,
141-147;201-284,
227-256;231-237,
292-375;318-347,
322-328;340-345,
383-503;419-435,
517-584;548-563,
574-602
127,215,308,453,
459,494
#disulfide_bonds #status predicted\
#binding_site carboxydrate (Asn) (covalent) #status
predicted\
#binding_site carboxydrate (Asn) (covalent) #status
experimental\
#active_site His, Asp, Ser #status predicted
SUMMARY #length 638 #molecular-weight 71273 #checksum 227
434,483,578
20.1%; Score 740; DB 1; Length 638;
Best Local Similarity 44.5%; Pred. No. 1,42e-142;
Matches 105; Conservative 45; Mismatches 78; Indels 8; Gaps 6;

Db 389 ARIWGTNSSLGEMPMQVLOVTVSONHMGSGSITGROMLTAAHCFD-GIYPDPVNR 446
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
253 SRIVGG-ESALPGMHPQVSLHVNQV---IVCGGSITTPETVITAAHCVKPLNNPHNWT 308

```

Db	Accession	Title	Organism	Date	Accessions	References	Authors	Journal	Title	Accession	Status	Molecule	Type	MRNA	Residues	Cross-References	Classification	Feature	Summary	Query Match	Best Local Similarity	Matches	Conservative	Score	DB 2	Length	Indels	Gaps	Checks	Sum	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB
----	-----------	-------	----------	------	------------	------------	---------	---------	-------	-----------	--------	----------	------	------	----------	------------------	----------------	---------	---------	-------------	-----------------------	---------	--------------	-------	------	--------	--------	------	--------	-----	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	----


```

388-625      388-625      #product coagulation factor Xla light chain #status
388-618      20-103,514-581,      #domain trypsin homology #label TRV\
571-599      46-76,50-56,      #disulfide_bonds #status predicted\
29      110-193,136-165,      #disulfide_bonds interchain #status experimental\
45      140-146,200-283,      #binding_site carboxylate (Asn) (covalent) #status
46      226-255,230-236,      #cleavage_site Arg-116 (coagulation factor Xla) #status
47      291-374,317-346,      #active_site His, Asp, Ser #status predicted\
48      321-327,380-500,      #binding_site carboxylate (Asn) (covalent) #status
49      416-432,545-560      #disulfide_bonds #status predicted\
90,126,353,450      #binding_site carboxylate (Asn) (covalent) #status

339      387-388      #disulfide_bonds interchain #status predicted\
#cleavage_site Arg-116 (coagulation factor Xla) #status
411,480,575      #active_site His, Asp, Ser #status predicted\
41      #binding_site carboxylate (Asn) (covalent) #status
#disulfide_bonds #status predicted\
#binding_site carboxylate (Asn) (covalent) #status

SUMMARY      #length 625      #molecular_weight 70109      #checksum 9314

Query Match      18.9%      Score 694: DB 1: Length 625:
      .Best Local Similarity 42.6%      Pred. No. 1,71e-131:
Matches 100: Conservative 43: Mismatches 86: Indels 6: Gaps 5:

Db 386 FRVGGTASVRGEMPVQVLTHTSPTRHILGGSIIIGMVLTAHCFY-GVESPKILRV 444
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 253 SRVVGESALPGAMPQVSLHVN-V-HVCGSIIIRPPIVVAHCEKPLNPMWHTA 309
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 445 YGILNQSEIKEDTSFGVQVEIIHQYKMAESGYDALLKLETTVYTSQRPICPSK 504
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 310 FGIIRFOSFMFYGAG-QYEKVISHPNYSKTRNNDAIMKRLKPLTFNDLVKPCLEPN 368
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 505 GGRNVIYTCWVWGAGYRKLRLDKIONTLOKAKIPLVNEGCKRYR-GKRTHKICAGY 563
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 369 GMLLPDQLCWLISGWCATEKEKTSSEYLNAAKVLITQGCNSHYVDNLITPMACIGF 428
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 564 REGGDADCKGDSGCPILSKCHNEFWMLVLTISMGEGCAQERPGVYTNVVEYVDNI 618
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 429 LQGNVDSGCGSGPLVTSKNNTIMWLTIGDPSWMSGCKAVRPYGVNVMFTDMI 483
      :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 9
ENTRY      S00845      #type complete
TITLE      hepsin (EC 3.4.21.-) - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
18-Jun-1999

ACCESSIONS S00845
REFERENCE   S00845
AUTHORS     Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie,
            E.W.
JOURNAL     Biochemistry (1988) 27:1067-1074
#title      A novel trypsin-like serine protease (hepsin) with a putative
            transmembrane domain expressed by human liver and hepatoma
            cells.
#cross-references MUID:88209431
#accession   S00845
#molecule_type mRNA
#residues    1-417 ##label LEY
#cross-references EMBL:X07732; NID:g32063; PIDN:CAA3058.1; PID:g32064

GENETICS    GDB:HPN: TMPRSS1: hepsin
#gene        ##cross-references GDB:135685; OMTM:142440
#map_position 19q11-19q13.2
CLASSIFICATION #superfamily hepsin; trypsin homology
FEATURE      hydroxylase; liver; serine proteinase; transmembrane protein

23-45      #domain transmembrane #status predicted #label TMN\
163-400      #domain trypsin homology #label TRY\
188-204,291-359,

```

Query Match	18.5%	Score 680	DB 1	Length 417
Best Local Similarity 33.4%		Pred. No. 3,96e-128		
Matches 120	Conservative	84	Mismatches 135	Indels 20
		Gaps 15		
322-338, 349-381	#disulfide_bonds	#status predicted		
203, 257, 353	#active_site His, Asp, Ser	#status predicted		
SUMMARY	#length 417	#molecular_weight 45011	#checksum 5762	
Query Match	18.5%	Score 680	DB 1	Length 417
Best Local Similarity 33.4%		Pred. No. 3,96e-128		
Matches 120	Conservative	84	Mismatches 135	Indels 20
		Gaps 15		
54	VOVSADARKLVFPDKTETWRLCLSSNSNAVAALSCSEEMGFALALHSELDVPTAAG 113			
149	VRLTGPNNILIVVSSQKRSWHPVQDDNNENYGAACRDGKKNFVSSO-GI-VDDS-G 205			
114	TSGFCEVDEGRLLPTQRLLEIVISVCDPRGRFLAICODCGRR-KLPVD-RIVGGRDTS 171			
206	STSFMKLNTSA-GNVDIKKTHSD-ACSKRAVSLRLCIACGVNLSNRQSLIVG-ESAL 262			
172	GRMPWOVSLRYDGAHLCCGSLSGDWVLTAAHCFPERNRVLSRWRYPAGAVADASP-HG 229			
263	PGAMPWOVSLHYQNVHCCGGSIITPEWITAAHCEKPERLNNPMTAFACILRQSFMYG 322			
230	LQLOVQAVYHGGVLPFPDRPSENSNDIALVHLSPLPLEYIQPCLPAAGALVDGK 289			
323	AGYQEVKISHPN-DSKTK-N-NDIALMKLQKPLTFENDVTKPVCLPFGMLDPEQ 376			
290	ICTVWGNPTGYQQAQGLVLEARPIISNDVCGADFYGNQIKPKFACGYPRGCGIDAC 349			
377	LCWISGMATEKERTSEVLNAAVLLIEIQRGMSRYVDLILPIAHICAGFLQGNVSC 436			
350	QDGSQGFVCEDSISRPRMRLCGVSKGTCALCAKQRYTKYSDPREWIFQAIKHS 408			
437	QDGSQGPLV--T-SKNINIMWLIGDTSWGSGLCAKAYRPGVYGVNMTPTDWIRQMRDQ 491			
RESULT 10				
TITLE	S33777	#type complete		
ORGANISM	hepsin (EC 3.4.21.-) - rat			
DATE	06-Jan-1995	#formal_name Rattus norvegicus	#common_name Norway rat	
ACCESSIONS	S33777	#sequence_revision 06-Jan-1995	#text_change 18-Jun-1999	
REFERENCE	S33777	S33777	S33777	
KEYWORDS	Barley, D.; Raymond, F.; Nick, H. Biochim. Biophys. Acta (1993) 1173:350-352			
FEATURE	#cloning_and_sequence_analysis of rat hepsin, a cell surface serine proteinase.			
22-44	#cross-references MIMD:93305733			
162-399	#accession S33777			
187-203, 290-358, 301-337, 348-380	#status Preliminary			
202,256,352	#molecule_type mRNA			
SUMMARY	#residues 1-416	#label FAR		
Query Match	17.8%	Score 652	DB 1	Length 416
Best Local Similarity 40.6%		Pred. No. 2.04e-121		
Matches 101	Conservative	52	Mismatches 83	Indels 13
		Gaps 9		
161	RIVGQDSSL-GRMPWOVSLRYDGTSLCCGSLSGDWVLTAAHCFPERNRVLSWRKRFAG 219			
254	RIVGG-ESALGAMPWOVSLHYQNVHCCGGSIITPEWITAAHCEKPERLNNPMTAFAG 312			
220	AVARTSP-HAVOLOVQAVYHGGVLPFPDRPISDNSNDIALVHLSPLPLEYIQPCLP 278			
313	ILRQSFMYGAGYQEVKISHPN-DSKTK-N-NDIALMKLQKPLTFENDVTKPVCLP 366			

#accession	A61545
#molecule-type	protein
#residues	1-33;34-117 #label SCH
REFERENCE	S17527
#authors	Schaller, J.; Straub, C.; Kaempfer, U.; Rickli, E.E.
#journal	Protein Seq. Data Anal. (1991) 4:69-74
#title	Complete amino acid sequence of equine mlnpIasminogen.
#cross-references	MUID:92052077
#accession	S17527
#molecule-type	proteIn
#residues	118-455 #label SC2
CLASSIFICATION	#superfamily plasmin: kringle homology; plasminogen-related protein precursor homology; trypsin homology fibrinolysis: glycoprotein; hydrolase; kringle; plasma; serine proteinase; zymogen
FEATURE	
1-33, 34-117,	#product plasminogen (fragments) #status experimental
118-455	#label PRO\
1-33	#domain activation peptide (fragment) #status experimental #label APV\
34-117,118-225,	#product plasmin (fragments) #status experimental #label MAT\
226-455	#domain kringle homology #label KR\
37-114	#product miniplasminogen #status experimental #label MIN\
118-455	#domain kringle homology #label KR2\
126-205	#domain plasmin chain B #status experimental #label BCH\
226-455	#domain trypsin homology #label TRY\
226-448	#active-site His, Asp, Ser #status predicted
267,310,405	#length 455 #checksum 1118
SUMMARY	
Query Match	16.8%; Score 617; DB 2; Length 455;
Best Local Similarity	38.2%; Pred. No. 4,65e-113;
Matches 97, Conservative	66; Mismatches 79; Indels 12; Gaps 12;
Db	210 PCCGPKVEPPKCSGVIGCVAINSWPQIGLRRFRGHFGCGTILSPBWYITAAHCL 269
OY	241 IACG-VNLNSSLR-QSRTIVGGESALPAMPQVSLHYO-NHVVGGSITPEWYITAAHCV 297
Db	270 ERS-SRPSTYKVVLGR-NHEELRLAANOQD-V-SKLELP-SRA-DIALKLSSPAIT 323
OY	298 EEPPLNPWHWTAFAGLLRDSFMYTGAGYOYEKVISHPNDYSKTKNNDIALKLOKPLEFN 357
Db	324 QNVIPACLPRADVVANMAECEFTVGTGETODSSNAG-VLKEADLPVIENVCN-RYEYLIN 381
OY	358 DLVKRVCLDPNGMLOPEDLCWISGNGATEEKTSVLANAVLLIEPTORCNSRYYY-D 416
Db	382 GRVKSTELCAGHLHVGVDSCQGDSGGPLVFCEKKYIILOGVTSMGLGCARPKNRGVYVR 441
OY	417 NLTTPAMICAGLGQCQSDSCQDGSGPLVTSKNIMWLIDGTSMGSCAKAYRPGYGYNV 476
Db	442 SSFIWMIERIKMSN 455
OY	477 WETDMTYRKMRAD 490
RESULT	13
ENTRY	P1UG
TITLE	#type fragment
ALTERNATE_NAMES	plasmin (EC 3.4.21.7) precursor - p1g (fragment)
CONTAINS	plasminogen
ORGANISM	mlnplasmaIngen
DATE	#formal_name Sus scrofa domestica #common_name domestic pig
ACCSSIONS	07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change
REFERENCE	18-Jul-1997
#authors	S03733; S03737; A2834
#journal	Schaller, J.; Martl, T.; Roesslelet, S.J.; Kaempfer, U.;
#title	Rickli, E.E.
	Fibrinolysis (1987) 1:91-102
	Amino acid sequence of the heavy chain of porcine plasmin.
	Comparison of the carbohydrate attachment sites with the

```
#accession      S03733          human and bovine species.
##molecule_type protein
#residues       1-560   #label SCH
REFERENCE
#authors        Brunisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.;
                    Lergler, M.; Manneberg, M.; Gillesen, D.
#journal         Eur. J. Biochem. (1981) 114:465-470
#title           Comparison of the primary structure of the N-terminal CNBR
                    fragments of human, bovine and porcine plasminogen.
#cross-references MUID:81212097
#accession      S03737
##molecule_type protein
#residues       1-57   #label BRU
REFERENCE
#authors        Martl, T.; Schaller, J.; Rickli, E.E.
#journal         Eur. J. Biochem. (1985) 149:279-285
#title           Determination of the complete amino-acid sequence of porcine
                    miniplasminogen.
#cross-references MUID:85203907
#accession      A25834
##molecule_type protein
#residues       450-790   #label MAR
FUNCTION
#description     dissolves the fibrin of blood clots; acts as a proteolytic
                    factor in a variety of processes including embryonic
                    development, tissue remodeling and tumor invasion; in
                    ovulation it weakens the walls of the graafian follicle;
                    also activates the urokinase-type plasminogen activator
                    fibrinolysis
#pathway          #superfamily plasmin: kringe homology; plasminogen-related
CLASSIFICATION  protein precursor homology; trypsin homology
                  fibrinolysis; glycoprotein; hydrolase; kidney; kringe;
KEYWORDS          plasma; serine proteinase
FEATURE
1-790             #product plasminogen #status predicted #label PRO\
                   (fragment) #label PLPH\
                   #domain activation peptide #status predicted #label APH\
                   #product plasmin chain A #status predicted #label ACH\
                   #domain kringe homology #label KR1\
                   #domain kringe homology #label KR2\
                   #domain kringe homology #label KR3\
                   #domain kringe homology #label KR4\
                   #product miniplasminogen #status experimental #label
                   MIN\
                   #domain kringe homology #label KR5\
                   #product plasmin chain B #status experimental #label
                   BCH\
561-790          #domain trypsin homology #label TRY\
561-783          30-54,34-42,84-162,
                   105-145,133-157,
                   166-243,169-297,
                   187-226,215-238,
                   256-333,277-316,
                   305-328,358-435,
                   379-418,407-430,
                   461-540,482-523,
                   511-535,547-665,
                   557-565,587-603,
                   679-746,709-725,
                   736-764
                   #disulfide_bonds #status predicted\
SUMMARY          #active_site His_Asp_Ser #status predicted
                  #length 790 #checksum 3770
QUERY MATCH      16.7%; Score 613; DB 1; Length 790;
Best Local Similarity 39.5%; Pred. No. 4,1de-112;
Matches 105; Conservative 52; Mismatches 94; Indels 15; Gaps 14;
```


QY 352 actaagaagcaactgtgcatcaacttgaccctggggaccttcctcgtggagctgcgctg 411
Db 241 actaagaagcaactgtgcatcaacttgaccctggggaccttcctcgtggagctgcgctg 300
QY 412 ggcgctgctactctctggaagttcattgagcagcaagtgtcccaactctggatagaatgc 471
Db 301 ggcgctgctactctctggaagttcattgagcagcaagtgtcccaactctggatagaatgc 360
QY 472 gacccctcagaatgacatcacaacccctcacaactggtgtgagtcggtgcacatgcgcc 531
Db 361 gacccctcagaatgacatcacaacccctcacaactggtgtgagtcggtgcacatgcgcc 420
QY 532 ggcggggagagcaggaatcgtgtgttcgcctctacagcacaacatcctacatgcagtg 591
Db 421 ggcggggagagcaggaatcgtgtgttcgcctctacagcacaacatcctacatgcagtg 480
QY 592 tactcatctcagaaggaagtcctgtgcaacctgtgtgcacaagaactggaacgaactac 651
Db 481 tactcatctcagaaggaagtcctgtgcaacctgtgtgcacaagaactggaacgaactac 540
QY 652 ggcggggagcagcctcagaagcattggtctataagaataatttctctagcacaagata 711
Db 541 ggcggggagcagcctcagaagcattggtctataagaataatttctctagcacaagata 600
QY 712 gtcgagtagcagcagatccaccagcttattgaaactgacaacaagtgcgcgcaatgctgat 771
Db 601 gtcgagtagcagcagatccaccagcttattgaaactgacaacaagtgcgcgcaatgctgat 660
QY 772 atctataaaaaactgtatcccaagtgatgctgtctctcaaaaagcagtgttctcttcgc 831
Db 661 atctataaaaaactgtatcccaagtgatgctgtctctcaaaaagcagtgttctcttcgc 720
QY 832 tgcataagccgaggggtgcaacttgaaactgaagccgcagagcagaagattgtggggcgag 891
Db 721 tgcataagccgaggggtgcaacttgaaactgaagccgcagagcagaagattgtggggcgag 780
QY 892 agcgcgtcccgaggggccttgccctgagcaggtcagcctgacagtcacgaacgttccagtg 951
Db 781 agcgcgtcccgaggggccttgccctgagcaggtcagcctgacagtcacgaacgttccagtg 840
QY 952 tgcgagaggtccatcatcaaccaccagtgatgctgtagacagccgcacgtcggtgaaaaa 1011
Db 841 tgcgagaggtccatcatcaaccaccagtgatgctgtagacagccgcacgtcggtgaaaaa 900
QY 1012 cctcttaacaatcatgctgagccttgagcagccttgaggggatttgagaaacattctcatg 1071
Db 901 cctcttaacaatcatgctgagccttgagcagccttgaggggatttgagaaacattctcatg 960
QY 1072 tctctatgagcagcatcaccagaagtagaaaaagtattctcatcccaattatgactccaag 1131
Db 961 tctctatgagcagcatcaccagaagtagaaaaagtattctcatcccaattatgactccaag 1020
QY 1132 accaagaacaatgacatgtgcgtgtgtagaagctgcagagaagccttgacttccaagacctta 1191
Db 1021 accaagaacaatgacatgtgcgtgtgtagaagctgcagagaagccttgacttccaagacctta 1080
QY 1192 gtgaagaacaaatgtgtcgtcccaaccaggatgattgtgtagcagcagaagaagctgtgtg 1251
Db 1081 gtgaagaacaaatgtgtcgtcccaaccaggatgattgtgtagcagcagaagaagctgtgtg 1140
QY 1252 atttcgaggtggtggggcagcaggaaggaagggagaaacctcagaagagtcggaagcgcgc 1311
Db 1141 atttcgaggtggtggggcagcaggaaggaagggagaaacctcagaagagtcggaagcgcgc 1200
QY 1312 aaggtgtctctcatctgagacacagagaatgcaacagcagatatgtctatgcaaacctgac 1371
Db 1201 aaggtgtctctcatctgagacacagagaatgcaacagcagatatgtctatgcaaacctgac 1280
QY 1372 acacacagcagatgctgtgctgcggtcttcctcaggggagaaagtcgattcttcgcaagggtgac 1431
Db 1261 acacacagcagatgctgtgctgcggtcttcctcaggggagaaagtcgattcttcgcaagggtgac 1320

QY 1432 agtggaggcctctgtgctactcgaagaacataatcgtgtgctgatataggatcaaacg 1491
Db 1321 agtggaggcctctgtgctactcgaagaacataatcgtgtgctgatataggatcaaacg 1380
QY 1492 tgggtgtctgtgctgtgccaagcttacaagccagggagtgtaaggggaatgtgagtattc 1551
Db 1381 tgggtgtctgtgctgtgccaagcttacaagccagggagtgtaaggggaatgtgagtattc 1440
QY 1552 acggactgattatcagacaatgagggcagacgctaa 1590
Db 1441 acggactgattatcagacaatgagggcagacgctaa 1479
RESULT 11
US-09-342-749-1
Sequence 1, Application US/09342749
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavtligian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMRSS2 Is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
FEATURE:
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. U75329.
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.

FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0227 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1077 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: SCORNOT01
 CLONE: 556016
 US-08-807-151-2

Query Match 57.1%; Score 992.4; DB 21; Length 1077;
 Best Local Similarity 99.8%; Pred. No. 3.6e-281;
 Matches 993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 721 agcgagatccagcgtttatgaactgaacacagtgccggcaatgcgatatctataa 780
 DB 83 AGCGAGTCCACGAGCTTTATGAACATGACCAAGTCCGGCAATGCGATATCTATAA 142
 QY 781 aaactgtacacagtgatgctgttctctaaagcagtggtttcttaactgtatagc 840
 DB 143 AAAGTGTACACAGAGTGTGCTGTCTTCAAAAGCAAGTGTCTTACCTGTATAGCC 202
 QY 841 tggggggtcaacttgaaactaaagccgagagagatgtggcgagagagcgctc 900
 DB 203 TGGGGGGTCAACTTGAACCTAACCCGACAGAGATCGTGGCGGCGAGAGCGCTC 262
 QY 901 cggggggtcgtgcttggcagtgacagctgcagctccagaacgctcagctgagagc 960
 DB 263 CCGGGGGCTGCGCTGGCGAGTCAAGCTCAGCTCCAGAACTCCAGTGTGCGGAGCG 322
 QY 961 tgcatacaccccgagtgatgctgtgacagcgcccgctgctggaaacctctaac 1020
 DB 323 TCCATATACCCCCGAGTGTGATGTCACAGCCGCTGCGTGAAGAAACCTCTTAC 382
 QY 1021 aatcctatgagcagcagcagcttgcggagattttgagacatcttccatctatga 1080
 DB 383 AATCATGCGATGAGAGCGATTTGCGGGATTTTGAGAAATCTTTCACTGTCTATAGA 442
 QY 1081 gccggaatacgaagtagaagaaagtattctcattcaataatactgaactcaagacaagac 1140
 DB 443 GCCGATACCAAGTAAAGATGATTCTCATCAAAATATGATCTCCAGACCAAGAAC 502
 QY 1141 aatgacattgctgtatgaagctgcgagaaagccctctgaacttcaagacctaagaaacca 1200
 DB 503 AATGACATTCGCTGTATGAAGCTGCAAGAACCTCTGCTTCAACGACCTAGTGAACCA 562
 QY 1201 gttgtctgcacacacacagcagatgctgcagccagacagcttgcgtatctccgg 1260
 DB 563 GTGTGCTGCCCAACCCAGGATGATGCTGCACCAAGACGCTGCTGATTTCCGG 622
 QY 1261 tggggggcagcagagagaaaggaagacctcagaagtgctgaaagctgcgcaagtgctt 1320
 DB 623 TGGGGGGCCACCCAGGAGAAAGGAAGACCTCAGAAAGTGTGAACGCTGCCAAGTGTCT 682
 QY 1321 ctcatgtgagacagagatgagcaacagatgtctatgacaaactctgatacaacagcc 1380
 DB 683 CTCTATTGAGACACAGAGATCAACACAGATATGCTATGACAACTGATCACACAGCC 742
 QY 1381 atgactgtgcagcgtctcctgcaggggaaagctcgatctctgcaggggtgacagtgagg 1440
 DB 743 ATGATGTGGCGGCTTCCGCGAGGGGAACGTGATTTCTGCCAGGGTGAACATGAGAGG 802
 QY 1441 cctctgtcaacttcgaagaacatactctgtgctgctaggggatacaagctggggtctc 1500

DB 803 CMTGTGCTCACTTGAAGAAATCTGTGTGCTGTATGAGGATACAAAGCTGGGCTTCT 862
 QY 1501 ggtctgcacaaagcttcaacagcagagtgatgacggaatgtatgtatctaacgagctg 1560
 DB 863 GGCTGTCCAAAGCTTACACACAGAGATGATGAGGATGATGATGATGATGATGATG 922
 QY 1561 attatcgacaatgagggcagagcggtctaaatcacaatgctctgctctgacgctgtt 1620
 DB 923 ATTATGACAAATGAGGGCAGAGCGCTATCCATAGTGTCTGCTGTGAGCTGTT 982
 QY 1621 tacaagaacacatggtggtgttcttcccccgtgacatgattctcttgaagatgat 1680
 DB 983 TACAGAAAACATGGGCTGTGTGCTTCCCGGTCATGATTTACTTTAGAGATGAT 1042
 QY 1681 tcagaggtcactcatctttatcaacagtgaaact 1715
 DB 1043 TCAGAGTCACTTATTATTTAAACAGTAACT 1077

RESULT 13

US-09-016-434-963
 Sequence 963, Application US/09016434

GENERAL INFORMATION:
 APPLICANT: Janice Au-Young
 APPLICANT: Jeffrey J. Sellhammer
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
 TITLE OF INVENTION: PATHWAY GENE EXPRESSION
 NUMBER OF SEQUENCES: 1490
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 PORTER DRIVE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/016,434
 FILING DATE: HERewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Zeller, Karen J.
 REGISTRATION NUMBER: 37,071
 REFERENCE/DOCKET NUMBER: PA-0002 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166
 INFORMATION FOR SEQ ID NO: 963:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1001 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: PROSTUT05
 CLONE: 842889
 US-09-016-434-963

Query Match 50.7%; Score 881.2; DB 34; Length 1001;
 Best Local Similarity 95.2%; Pred. No. 2e-248;
 Matches 936; Conservative 9; Mismatches 31; Indels 7; Gaps 4;

QY 21 gcgagggcgggagcgccgctgga--gcgcgagagtcataatgacattcagabata 77

```
; EARLIER APPLICATION NUMBER: 60/125,469
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 60/125,560
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 1296
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1044
; LENGTH: 1181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1181)
; OTHER INFORMATION: n = A,T,C or G
US-09-526-996-1044
```

```
Query Match      26.7%; Score 463.6; DB 92; Length 1181;
Best Local Similarity 98.9%; Pred. No. 2,2e-125;
Matches 466; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 481 ggtaccctgacataccctctaaactgtgtgatgggtgtcaacatgcccgcggggag 540
Db 711 ggtaccctgacataccctctaaactgtgtgatgggtgtcaacatgcccgcggggag 770
QY 541 gacgagaatcgtgtgtgttcgctcttaagacgaaacttcaccttcaggttactatct 600
Db 771 gacgagaatcgtgtgtgttcgctcttaagacgaaacttcaccttcaggttactatct 830
QY 601 cagagagaagtccttggaacctgtgtgccaagaagactggaacgaaactacggggggcg 660
Db 831 cagagagaagtccttggaacctgtgtgccaagaagactggaacgaaactacggggggcg 890
QY 661 gacctgacggagacataggagataattttactcttagccaaggaaatagtgatgac 720
Db 891 gacctgacggagacataggagataattttactcttagccaaggaaatagtgatgac 950
QY 721 agcgatccacacagcttatagaactgaaacaaagtgcggcaatgtcgaatatataaa 780
Db 951 agcgatccacacagcttatagaactgaaacaaagtgcggcaatgtcgaatatataaa 1010
QY 781 aaactgtacacacagtgatgctgttcttcaaaagcaatgttcttcaagctgtatagcc 840
Db 1011 aaactgtacacacagtgatgctgttcttcaaaagcaatgttcttcaagctgtatagcc 1070
QY 841 tggggggtacactgaactgaacccgacagagagatgttgggggcggaagagcgcgctc 900
Db 1071 tggggggtacactgaactgaacccgacagagagatgttgggggcggaagagcgcgctc 1130
QY 901 ccgggggcttgcccttgcccttgcccttgcccttgcccttgcccttgcccttgcccttg 951
Db 1131 ccgggggcttgcccttgcccttgcccttgcccttgcccttgcccttgcccttgcccttg 1181
```

Search completed: April 22, 2000, 06:15:46
Job time: 2285 sec

Db 417 AACCAAGTGTCTCTCCACCAACGACCAATGATCTGACCCAGAACAGCTCTGCTGATT 358
 Oy 1255 tccgggtggtggccaccgagagaaagaaagacccagagctgctgaagctgccaag 1314
 Db 357 TCGGGGTGGGGGCCACCGAGAGAGAAAGGAAAGACTCAAAAGTGTGAACGCTCCAAAG 298
 Oy 1315 gtccttctcatgagacacagagatgcaacagcaagatgctatgacaacctgataca 1374
 Db 297 GTGCTTCTCATGAGACACAGAGATGCAACAGCATATGTCTATGACAACTGATCACA 238
 Oy 1375 ccagcagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1434
 Db 237 CCAGCCATGATCTGTGCTCCGCTTCTGACAGGGAACGTCGATTCCTGCAAGGCTGAC 178
 Oy 1435 ggaaggccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1494
 Db 177 GAGAGCCCTGCTGCTCCCTTCGAAAGAAATATCTGCTGCTGATAGGGATACAACTGG 118
 Oy 1495 ggtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1554
 Db 117 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 58
 Oy 1555 gactgattatcgacaatgagagagagagagagagagagagagagagagagagagag 1611
 Db 57 GACTGATTATTCGACAAATGAGAGGACAGGCTAATCCATGATGTTTCTGCTCTG 1

RESULT 4
 A1394553 385 bp mRNA EST 04-FEB-1999
 LOCUS A1394553.1 NC1-CGAP-CL1 Homo sapiens cDNA clone IMAGE:2108144 3'
 DEFINITION similar to TR:015393 SERINE PROTEASE.; mRNA sequence.
 ACCESSION A1394553
 VERSION A1394553.1 GI:4224100
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiinae; Homo.
 1 (bases 1 to 385)
 NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Mar 9, 1998 this sequence version replaced gi:2946770.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
 M.D., Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NC1-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html

FEATURES
 source
 Trace considered overall poor quality
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1. 385
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="NC1-CGAP-CL1"
 /clone_lib="NC1-CGAP-CL1"
 /tissue_type="B-cell, chronic lymphocytic leukemia"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I. Site 2: Eco RI. 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTCACCAATCTGAGGTGGAGCGCGCATTCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
 T 3'] double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified p773 vector.
 Library is normalized, and was constructed by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 82 a 116 c 85 g 102 t
 ORIGIN

Query Match 21.3%; Score 370.6; DB 45; Length 385;
 Best Local Similarity 97.7%; Pred. No. 5,2e-91;
 Matches 376; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 1201 gtcgtctcccaaccagcagatgctgctgctgctgctgctgctgctgctgctgctgctg 1260
 Db 385 GTGTGCTCTCCCAACCCAGGATTAATGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 326
 Oy 1261 tggggggccaccgagagagagagagagagagagagagagagagagagagagagagag 1320
 Db 325 TGGGGGGCCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 266
 Oy 1321 ctcattgagacacagagatgcaacagagatgctatgacaacctgatacaccagcc 1380
 Db 265 CTCATTAAAGACACAGAGATGCAACACAGATATGCTATGACAACTGATCACACAGCC 206
 Oy 1381 atgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1440
 Db 205 ATGATTTGTGCGCGCTTCTGCAAGGGAACGTCGATTTCTGCAAGGGAACGTCGAGGG 146
 Oy 1441 cctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1500
 Db 145 CCTCTGCTCACTTCGAAAGAAATATTTGCTGCTATAGGGGATACAAACTGGGTTCT 86
 Oy 1501 ggcctgcaaaagcttaccagacagagagctgctgctgctgctgctgctgctgctgctg 1560
 Db 85 GGCTGTCGCAAAAGCTTACAGACAGAGAGTACGGGAATGTGATGATTCACAGGACTGG 26
 Oy 1561 attatcgacaatgagggcagacg 1585
 Db 25 ATTTATCGACAAATGAGGGCGACG 1

RESULT 5
 A1261741/c 353 bp mRNA EST 03-FEB-1999
 LOCUS A1261741.1 NC1-CGAP-1411 Homo sapiens cDNA clone IMAGE:2028487 3'
 DEFINITION similar to TR:015393 SERINE PROTEASE.; mRNA sequence.
 ACCESSION A1261741
 VERSION A1261741.1 GI:3869944
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiinae; Homo.
 1 (bases 1 to 353)
 NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jan 14, 1998 this sequence version replaced gi:1797896.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NC1-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html

Insert Length: 271 Std Error: 0.00
 Seq primer: -40UP from Gibco

DB	145	7GTCCAGACTTCCTTGTGCTCAACACACTTCTGCAGAAAGAACCAAGGCGCTGAATTT	86
OY	1647	gcttcccgctgcagatctactcttagagatgattcagaggtcacttatttataaa	1706
DB	85	AACCTCTGCTGACAAATGTACCTTTGAGATGATTCGAAGGCGCTTTCACCTTTATTA	26
OY	1707	cagttg	1712
DB	25	cagta	20

RESULT	7
AI480993/C	564 bp mRNA EST 09-MAR-1999
DEFINITION	v166909.x1 Soares mouse mammary gland NbMWG Mus musculus cDNA clone
IMAGE	850720 3' similar to TR:O15393 O15393 SERINE PROTEASE. ;
RNA sequence	
AI480993	
VERSION	AI480993.1 GI:4374219
KEYWORDS	EST.
ORGANISM	house mouse.
SOURCE	Mus musculus
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia; Eutheria: Rodentia; Sclurognathi, Muridae; Murine; Mus.
AUTHORS	1 (bases 1 to 564) Matta,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ralfer,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Wilson,R.
TITLE	The WashU/NCI Mouse EST Project 1999
JOURNAL	Unpublished (1999)
COMMENT	On Jun 22, 1998 this sequence version replaced g1:3247039. Contact: Marra M/WashU/NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG:502872 This clone was previously sequenced on the 5' end only, this new data is from the 3' end Possible reversed clone: similarity on wrong strand High quality sequence stop: 353. Location/Qualifiers 1..564 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:850720" /clone_lib="Soares mouse mammary gland NbMWG" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B" /note="Organ: mammary gland; Vector: p773D-Pac (pMacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5]. TGTTACCAATCTGAAGTGGAGCGCGCGAGATGTTTTTTTTTTTTTTTTTTT T 3 1; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

FEATURES	SOURCE
BASE COUNT	135 a 150 c 129 g 150 t

Query Match	17.7%	Score 307.2	DB 47	Length 564
-------------	-------	-------------	-------	------------

[illegible]

```

OY 308 TAFAGILNOSFM-FYGAQYQYEVKISHPNYDSKTKNNDIAMKLOKPLTFNDLVKPCVLP 366
DB 122 GPGPPLATGTCWVNGLVHSGEALAS-VLQEVAVPLDSDNMCELMYLHGEPSLAGORLI 180
OY 367 NPGMMLQPEQLCWMISGWCATEKGTSEVLAANKVLLIETORCNSRYV-D-NL-----I 419
DB 181 QDDMLCAGSVGKRDSCGDSGGLPCPINDTWTIOAGIVSNMGFCARFPRGVYQVLSY 240
OY 420 TPAMICAGFLGNDVSCGDSGGLPVTSKNNIMWLIGDTSNGSGCAKAYRGVGNVWF 479
DB 241 TDMTQRTL 248
OY 480 TDMTQRTM 487

RESULT 13
ID W77302 standard; Protein; 271 AA.
AC W77302;
DE Amino acid sequence of SP001LA, a homologue of HELA2.
KW Serine protease; regulation; cell activity; viability; HELA2; ATC2;
KW BOM3; testis; fertility; suppressor; testicular germ cell cancer;
KW seminoma; testis-specific expression; antitumour; sperm development;
KW infertility; human; chromosome 16p13.3.
OS Homo sapiens.
FH Key
FS Disulfide_bond 29 Location/Qualifiers
FT /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be a catalytic residue"
FT Disulfide_bond 45 /note- "likely to be a catalytic residue"
FT /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be a catalytic residue"
FT Disulfide_bond 115 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 129 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 160 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 162 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 181 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 192 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 196 /note- "likely to be a catalytic residue"
FT /note- "likely to be a catalytic residue"
FT Disulfide_bond 202 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 220 /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be involved in disulphide bonding"
FT /note- "likely to be a catalytic residue"
PN W09836054-A1.
PD 20-AUG-1998.
PR 13-FEB-1998; AU0085.
PR 18-NOV-1997; AU-000422.
PR 13-FEB-1997; AU-005101.
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
PI Antalis TM, Hooper JD;
DR WPI: 98-480768/41.
DR N-PSDB: V59134.
PT New serine protease(s) and kinase involved in regulating cell
PT activity and viability - particularly the testis-specific protease
PT HELA2 used for modulation of fertility and as tumour suppressor
PS Example 15; Fig 20A; 167pp; English.
CC W77302-04 represent HELA2 homologues. The genes are found in a cluster
CC on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
CC homology to serine proteases. The protein is involved in or associated
CC with regulation of cell activity and/or viability. Administration of
CC recombinant HELA2 (also called testis) is used to increase fertility.

```

```

CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
CC testicular germ cell cancers (seminoma) and is also expressed in some
CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is
CC a marker/potential therapeutic target for cancer. The promoter from the
CC HELA2 gene is useful for testis-specific expression of other genes,
CC e.g. for gene therapy or modulation of fertility. Drugs that block
CC activity of HELA2 should have antitumour activity (other than in
CC testis) while in testis recombinant HELA2 should stop growth of tumours
CC and normalise sperm development (eliminating the need for orchidectomy).
CC Identification of mutant forms of HELA2 can be used to diagnose
CC infertility.
SO Sequence 271 AA;

Query Match 16.6%; Score 608; DB 1; Length 271;
Best Local Similarity 37.4%; Pred. No. 5,806-48;
Matches 89; Conservative 54; Mismatches 88; Indels 7; Gaps 5;

DB 2 NRVYGEDSTSEMPWYSIOKNGTHHCAGSLTSRWVITAHACEKDNLPFLSVLG 61
OY 253 SRVGGESALPGAMPQVSLHVOVNVHCGSITPEWIVTAHCVEKPLNPMWHTAFAG 312
DB 62 AMQGNPESRSQKQVAVPEHPHYVSWKEGACADIALVRLERSIOFSRVLPICLPDAI 121
OY 313 ILROSEMFY-GAGYQVEKIVISHPNYDSKTKN-NDIALKLOKPLTFNDLVKPCVLP 370
DB 122 HLPPTHGCMISGMSIOQVPLPHPOTLQKLKLPIDSEVCSHLVWRGAGCGPIEDMIC 181
OY 371 MLOPEQLCWMISGWCATEKGTSEVLAANKVLLIETORCNSRYV-D-NL-----I 419
DB 182 AGYLEGERDACLDGSGGLPCVMDGAWLLAGIISWEGCARNRPGVYISLSAHRSV 239
OY 426 AGFLGNDVSCGDSGGLPVTSKNNIMWLIGDTSNGSGCAKAYRGVGNVWF 483

RESULT 14
ID Y13391 standard; Protein; 317 AA.
AC Y13391;
DE Amino acid sequence of protein PRO343.
KW Secreted protein; transmembrane protein; human; enterocolitis;
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW congenital microvillus atrophy; skin disease; cell growth;
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophla areata;
KW anti-Chromoblastic; wound healing; tissue repair.
OS Homo sapiens.
FH Key
FS Homo sapiens.
PN W09914328-A2.
PD 25-MAR-1999.
PR 16-SEP-1998; U19330.
PR 17-SEP-1997; US-056840.
PR 17-SEP-1997; US-059113.
PR 17-SEP-1997; US-059115.
PR 17-SEP-1997; US-059117.
PR 17-SEP-1997; US-059119.
PR 17-SEP-1997; US-059121.
PR 17-SEP-1997; US-059122.
PR 17-SEP-1997; US-059184.
PR 18-SEP-1997; US-059263.
PR 18-SEP-1997; US-059266.
PR 15-SEP-1997; US-062125.
PR 15-SEP-1997; US-062126.
PR 17-SEP-1997; US-062285.
PR 17-SEP-1997; US-062287.
PR 21-OCT-1997; US-063486.
PR 21-OCT-1997; US-062814.
PR 24-OCT-1997; US-062816.
PR 24-OCT-1997; US-063045.
PR 24-OCT-1997; US-063120.
PR 24-OCT-1997; US-063121.
PR 24-OCT-1997; US-063122.
PR 24-OCT-1997; US-063127.
PR 24-OCT-1997; US-063128.
PR 27-OCT-1997; US-063329.
PR 27-OCT-1997; US-063327.

```


Search completed: Wed Apr 19 23:28:04 2000
Job time : 312 secs.

```

OY 113 CSNSGIECCSSGTCINPSMWCDSVSHCPGEGEDNRCVRLY-GP-NFI-LQVYSSQKRSW 169
DB 480 VACAEWMTQISDDVQGLGLGTG-NSVPTFS-TGGGPVINTAP-NSLLLT-PSQ 535
OY 170 PVCQDDMNENYGAARCRDGYKNNEYSSQIYDDSGTSTPMKINTSAGNDYIKKTHSDA 229
DB 536 CLEDSILLQCNKSCGKKLVTOEVPKTYGSDSREGAMPVVAAYFDQOYCGASLVS 595
OY 230 CSKAVVSLRCT--ACGVNLNSROS-RIVGGSALPGAMPQVSLHVQVHVHCGGSIIT 286
DB 596 RMLVSAHCYVGRNME-PSKMAVLGLHMASLSPQIETRLIDQIVINPHYKRRKN 654
OY 287 PEMIVAAHCY-EKPLNPMHMTAFAGILLQSPMEFYGA-GYQ-VEKYISHPNYDSKTKN 343
DB 655 DIAMHLEKVNITDYIOPICLPEENOVFPFGRCISAGKALITGSDTADVQEDVPL 714
OY 344 DIAMLRQKRLFTFNDLVKPCVLPNPGMILQPEQLCWSIGWATEEKGKSEVLNAKAVLL 403
DB 715 LSNKCGCOQ-MPEYNITENNVACAGYAGVSDCGSDSGPGLMCQENNRMLLAGVTSFGYQ 773
OY 404 IETORNSRKYVNLITPMICAGFLQGNVSDCGSDSGPLVSKNNIMWLIDTISWGS 463
DB 774 CALPNRPYAVRVPRTETWI 793
OY 464 CAKAYRPGVGNVMTFTDWI 483

```

```

RESULT 2
ID W22986 standard; Protein: 233 AA.
AC W22986;
DT 08-OCT-1997 (first entry)
DE Human serine protease 60 (SP60).
KM Human; colon carcinoma, COLO 201; cell line; serine protease; SP60;
KW screening; inhibitor; treatment; disease.
OS Homo sapiens.
PN J09149790-A.
PD 10-JUN-1997.
PF 24-JUL-1996; 212196.
PR 29-SEP-1995; JP-275105.
PA (SUNR) SUNTORI LTD.
DR MPI: 97-357902/33.
N-PSDB: T79127.
PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
PT - useful to screen for specific inhibitors, e.g. to search for, or
PT study agent for treatment of various diseases
PS Claim 1; Pages 11-12; 16pp; Japanese.
CC The present sequence is the human colon carcinoma COLO 201
CC cell line derived serine protease 60 (SP60), which can be used to
CC screen for specific inhibitors, e.g. to search for, or study an
CC agent for the treatment of various diseases.
SQ Sequence 233 AA;

```

```

Query Match 19.3%; Score 710; DB 1; Length 233;
Best local Similarity 44.6%; Pred. No. 2,36e-58;
Matches 104; Conservative 48; Mismatches 71; Indels 10; Gaps 9;

```

```

DB 1 VVGGEASVDSMPQVSIQYDKQVCGSILDRHVVITAHCFKXK-TDVPNKKVYRAGSD 59
OY 255 IYGGESALPGAMPQVSLHVQVHVHCGGSIITPEMIVTAACHVEKPLNNPMHMTAFAGIL 314
DB 60 KLG-SF--PSLAIAKII--TEFNPMYKNDIALMKLOPLTFSGVTRPCLPFEDEL 115
OY 315 ROSFMFGAGYQVEKYVSHRNYDSK-TKNDIALMKLOKRLTFDVLKPCVLPNPGMILQ 373
DB 116 PATPL-WIIGMFTKONGKMSDILLQASVOVIDSTRCNADAVGEGTEKMKAGIPEG 174
OY 374 PE-QLCHISMGATEEG-KTSEVLNAKAVLLIETORNSRKYVNLITPMICAGFLQ 431
DB 175 GVDTGSDSGPLMYSQD-WHVVGIYSKYGCGGSPSTPGYTVVSAVLWKIY 226
OY 432 NVDSQGDSDGSPVLTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVMTFTDWI 484

```

```

RESULT 3
ID R89435 standard; Protein: 418 AA.
AC R89435;
DT 27-MAY-1996 (first entry)
DE Trypsin-like enzyme.
KM Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW protease; primer; PCR; amplification.
OS Homo sapiens.
FH Key
FT Peptide 1..186
FT Protein /label_sig_peptide
FT /label_mat_protein
PN AU9527248-A.
PD 08-FEB-1996.
PF 31-JUL-1995; 027248.
PR 29-JUL-1994; JP-178607.
PA (TEIJ) TEIJIN LTD.
PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;
PI Yamoka K, Yasuoka S;
DR MPI: 96-117356/13.
DR N-PSDB: T10689.
PT Nucleic acid sequence encoding trypsin-like enzyme - which digests
PT fibrinogen, used as expectorant in treatment of respiratory
PT diseases, e.g. bronchial asthma
PS Example 11; Page 47-49; 65pp; English.
CC The overlap parts of 107 bp between the sequences given in
CC T10698 and T10703 were identified, and thereby their identity
CC was confirmed. It was confirmed from the sequence analysis,
CC that these overlapping sequences contained a region encoding the
CC amino acids of the N-terminus 20 residues of the trypsin-like
CC enzyme isolated from the cough phlegm. The sequences were
CC ligated, and the desired trypsin-like enzyme gene cDNA sequence
CC was determined (T10689).
SQ Sequence 418 AA;

```

```

Query Match 19.0%; Score 697; DB 1; Length 418;
Best local Similarity 42.3%; Pred. No. 5.04e-57;
Matches 105; Conservative 49; Mismatches 87; Indels 7; Gaps 6;

```

```

DB 173 CGA-GPDLITLSEORILGTEAEKSGMPQVSLRLNNAHCGSLINNMILTAHCF-R 230
OY 240 CINCQVNLNSRQSRIVGSGALPGAMPQVSLHVQVHVHCGGSIITPEMIVTAACHVEK 299
DB 231 SNSNPRDMITSGI-STTP--KLRRKVRNILLNNYKSTHENDALVRLNSVFTFD 287
OY 300 PLNNPMHMTAFAGILRQSEMFYAGYQVEKYVISHPNYDSKTKNDIALMKLOKPLTFNDL 359
DB 288 IHSVCPLPAATQNIIPGSTAVVTGMAQOEYAGTVPELRQGOVRIISNDVCNAPHSYNGAI 347
OY 360 VKPVCCLNPGMILQPEQLCHISMGATEEKGKSEVLNAKAVLLIETORNSRKYVDNLI 419
DB 348 LSGMLCAGVPGGVDAOCGSDSGPLVQEDSRRLMFTVGIVSMGDCGLPDKPGVYTRVTA 407
OY 420 TPAMICAGFLQGNVSDCGSDSGPLVTSKNN-IMWLIGDTSWGSCKAKAYRPGVGNVAV 478
DB 408 YLDMT-RQ 414
OY 479 FTDIYRQ 486

```

```

RESULT 4
ID W46917 standard; Peptide: 356 AA.
AC W46917;
DT 02-JUL-1998 (first entry)
DE Amino acid sequence of a novel human kallikrein.
KM Kallikrein; HKLP; human; serine protease; drug screening; atagonist;
KW agonist; treatment; hypertension; cardiac hypertrophy; arthritis;
KW inflammatory disorder; blood clotting disorder.
OS Homo sapiens.
FH Key
Location/Qualifiers

```

```

DEFINITION      tg09d11.x1 NCI-CGAP-CLL1 Homo sapiens cDNA clone IMAGE:210827 3'
ACCESSION       AJ393270
VERSION         AJ393270.1 GI:4222817
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE       1 (bases 1 to 450)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL         Unpublished (1997)
COMMENT         On Jan 19, 1998 this sequence version replaced gi:2153106.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrrp/image/image.html

FEATURES
   source
       1..450
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="IMAGE:2108277"
           /clone_1ID="NCI-CGAP-CLL1"
           /tissue_type="B-cell, chronic lymphocytic leukemia"
           /lab_host="DH10B"
           /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
           polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
           was primed with a Not I - oligo(dT) primer [5',
           TGTTCACCAATCTGAAGTGGGCGCGCGATGCTTTTATTTTATTTTATTTT
           T 3']; double-stranded cDNA was ligated to Eco RI
           adaptors (Pharmacia), digested with Not I and cloned into
           the Not I and Eco RI sites of the modified pT7T3 vector.
           Library is normalized, and was constructed by Bento
           Soares and M. Fatima Bonaldo."

BASE COUNT      98 a      132 c      107 g      111 t      2 others
ORIGIN
Query Match      25.7%; Score 446.4; DB 45; Length 450;
Best Local Similarity 99.3%; Pred. No. 9.1e-112;
Matches 447; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1162 ctgcagaagctctgaccttaagaactagtgaaacagtgctgctccaaaccaggc 1221
    |||||||
Db 450 CTGCAGAAGCTCTGACCTTTCANCACTAGTGAACACAGTGTCTGCCCAACCCAGGC 391

Oy 1222 atggtctctgcagcagaacagctctgctggtttccgggttg99ggccacagagaaga 1281
    |||||||
Db 390 ATGATGCTGCAGCCGACGACAGCTCTGCTGNNATTTCCGGGTGGGGGCCACGAGAGAAA 331

Oy 1282 gggaagacctcagaagtgctgaacgctgcgaagtgcttcctcaatgagacagagatgc 1341
    |||||||
Db 330 GGGAAGACCTCAGAAGTCTGTAACGCTGCCAAGGTGCTTCATATGAGACACAGAGATGC 271

Oy 1342 aacggcagatatgtctatgacaacctgtatcacaccagcagatgtggcggctctctg 1401
    |||||||
Db 270 AACGACAGATATGTCTATGACAACTGATCACACGACGATGATGTGGCGGCTTCCG 211

Oy 1402 cagggaacgctcgatctctgcagggtcacagtgaggggccctgtcaactcagaagac 1461
    |||||||

```

```

Db 210 CAGGGGAACGTGATCTTGTCCAGGCTGACAGTGGAGGCCCTCTGCTACTTCGAGAAC 151
Oy 1462 aatctgtgtgcgtatagaggatatacaagctcg99gtctgctgctgcacaaagctacaga 1521
    |||||||
Db 150 AATATCTGTGCTGCTATATGGGGATTCAGAGCTGGGCTTGTGCTGTGCCAAAGCTTACGA 91

Oy 1522 ccaggagtgctacgggaatgtatgtatcttcacggactgagattatcgaacaaaggagga 1581
    |||||||
Db 90 CCAGAGAGTACGGGAAGATGATGATATTCACGCACTGATTTATCGCAATGAGCGCA 31

Oy 1582 gacggtatccacacatggtcttcgctctg 1611
    |||||||
Db 30 GACGCTAATCCACATGATGCTTCTCTTG 1

RESULT          3
LOCUS           A1673506 417 bp. mRNA EST 18-MAY-1999
DEFINITION     w19h04.x1 Soares_Dieckgraefe_colon_NHUC Homo sapiens cDNA clone
IMAGE:2351095 3' similar to SW:TMS2_HUMAN O15393 TRANSMEMBRANE
PROTEASE, SERINE 2 ;, mRNA sequence.
ACCESSION      A1673506
VERSION        A1673506.1 GI:4853237
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE       1 (bases 1 to 417)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL         Unpublished (1997)
COMMENT         On May 18, 1998 this sequence version replaced gi:3138221.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from g1bco
High quality sequence stop: 215.
Location/Qualifiers
    1..417
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2351095"
        /clone_1ID="Soares_Dieckgraefe_colon_NHUC"
        /tissue_type="colonial mucosa from 5 ulcerative colitis
        patients"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
        modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st
        strand cDNA was primed with a Not I - oligo(dT) primer [5',
        TGTTCACCAATCTGAAGTGGGCGCGCGATGCTTTTATTTTATTTTATTTT
        T 3']; double-stranded cDNA was ligated to Eco RI adaptors
        (Pharmacia), digested with Not I and cloned into the Not I
        and Eco RI sites of the modified pT7T3 vector. Library
        went through one round of normalization. Tissue samples
        provided by Dr. Brian Dieckgraefe (Washington University,
        dieck@im.wustl.edu); colonic mucosa represents a range of
        disease involvement from mild cryptitis to severe
        ulceration, fibrosis, and degeneration. Library
        constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      90 a      125 c      98 g      103 t      1 others
ORIGIN
Query Match      22.7%; Score 395.2; DB 50; Length 417;
Best Local Similarity 96.6%; Pred. No. 9.6e-98;
Matches 403; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 1135 aaacagtgctctgcacacccagcagcatgatgtcgcagcagaacagctctgctgatt 1234
    |||||||

```

CC and are useful to determine effective treatment for bone and skeletal
 CC disorders such as osteoporosis, Paget's disease and osteitis
 CC deformans, especially those associated with elevated bone alkaline
 CC phosphatase levels.
 SQ Sequence 416 AA;

Query Match 18.2%; Score 668; DB 1; Length 416;

Best Local Similarity 33.7%; Pred. No. 4,59e-54;
 Matches 121; Conservative 77; Mismatches 141; Indels 20; Gaps 15;

DB 53 VQSPDRLAVDTKEGTRRLCSRNARVAGLCEEGFRLALAHSLDVTGTGANG 112
 QY 149 VRLYGNFLQVYSSORRSHHPYCGODMNMENYRAACRDMGYKNNYSSOGYVDSGSTS 208
 DB 113 TSGFPCVDEGSLRLAORLLDVISVCDPCPRGRFLATCCDGR-R-KLPVD-RIVGGDSSL 170
 QY 209 FMKLN-TSAGNDVYIKRY-HS-DACS-SKA-VYSLRCIACGVNLNSRGRITYG-ESAL 262
 DB 171 -GMPQVSLRLDGTLLCGGSLSGMVLTAHCFPERNRVLSRMVAFAGAVARTSP-NA 228
 QY 263 PGAMPQVSLHVNQNVHVCSSITTPMIVTAHCEKPLNNPWHMTAFAGILRQSEMEFYG 322
 DB 229 VQGVQAVIYHGYLFPRTPTDENNDIALVHLSSSLPTEYIOVCLPAAGALVDGR 288
 QY 323 AGYQVKEVISHPNY---DSKTR-N-NDIALMLKLPKPLTFNDLVKVCPLPFGMDPEQ 376
 DB 289 VCTVTGNGWTFQYGOQAMVLOEARVPIISNENYSPDEFYGNQIKPMKFCAGYEGGIDAC 348
 QY 377 LCVISGKATEEGKTSYEVLNAAKVLLEIQRNSRYVDNLTPMICAGFLQGVNDSC 436
 DB 349 CGDSGPEVCEDSISGTSNRRLCGIVSWGTGALARRPGVYTVTDREMIFAIKTHS 407
 QY 437 CGDSGPEV--TSKNNI--WMLIGDTSWGSCKAKAYRPGVYGVNMTDMYIQRMDAG 491

RESULT 7
 ID W77301 standard; Protein: 285 AA.

AC W77301;
 DT 07-JAN-1999 (first entry)
 DE Amino acid sequence of murine HELA2.
 KM Serine protease; regulation; cell activity; viability; HELA2; ATC2;
 KM BCM03; testisin; fertility; suppressor; testicular germ cell cancer;
 KM seminoma; testis-specific expression; antitumor; sperm development;
 KM infertility; mouse.
 OS Mus sp.
 PN WO9836054-A1.
 PD 20-AUG-1998.
 PF 13-FEB-1998; AU0085.
 PR 18-NOV-1997; AU-000422.
 PR 13-FEB-1997; AU-005101.
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 PI Antalis TM, Hooper JD;
 PI WPI: 98-480768/41.
 DR N-PSDB; V59132.
 DR New serine protease(s) and kinase involved in regulating cell
 PT activity and viability - particularly the testis-specific protease
 PT HELA2 used for modulation of fertility and as tumour suppressor
 PS Example 14; Fig 18A; 167pp; English.
 CC The present sequence represents the amino acid sequence of murine HELA2.
 CC Human HELA2 was isolated from Hela cells. HELA2 has high homology to
 CC serine proteases. The protein is involved in or associated with
 CC regulation of cell activity and/or viability. Administration of
 CC recombinant HELA2 (also called testisin) is used to increase fertility.
 CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
 CC testicular germ cell cancers (seminoma) and is also expressed in some
 CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is
 CC a marker/potential therapeutic target for cancer. The promoter from the
 CC HELA2 gene is useful for testis-specific expression of other genes,
 CC e.g. for gene therapy or modulation of fertility. Drugs that block
 CC activity of HELA2 should have antitumor activity (other than in
 CC testis) while in testis recombinant HELA2 should stop growth of tumours
 CC and normalise sperm development (eliminating the need for orchidectomy).
 CC Identification of mutant forms of HELA2 can be used to diagnose

CC infertility.
 SQ Sequence 285 AA;

Query Match 17.8%; Score 653; DB 1; Length 285;

Best Local Similarity 39.0%; Pred. No. 1.55e-52;
 Matches 97; Conservative 53; Mismatches 87; Indels 12; Gaps 10;

DB 14 SRVGGDDAEICRMWQSGSLRWGNHLCATLLNRRVITAAHCRKD-NDPFDVTOFG 72
 QY 253 SRVGGESALPGAMPQVSLHVNQNVHVCSSITTPMIVTAHCEKPLNNPWHMTA-FA 311
 DB 73 ELTSRSLNNLQAYSNRVOIEDIFLSPKSEBY-PNDIALKLSSPYNNIOPICLN 131
 QY 312 GIL-RQSF--M-FYAGYQVKEVISHPNYDTSKKNNDIALMLKLPKPLTFNDLVKVCPLN 367
 DB 132 STYKFNRRDQVWTGGAIGEDSLPSPTLQEVQVAIINNSCMHMYKPPDFTINMD 191
 QY 368 PGMMLOPEQLCHISWGAT-EKRG-KTSEVLAARVLLIETORCSRY--V-YDMLITPA 422
 DB 192 MYCAGTPEGKDACFEGSGPLACDDDTWYQGVVSWGICGGRNRPQVYTNISHYNN 251
 QY 423 MICAGFLQGVNDSCQDSGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNMTDM 482
 DB 252 IOSTMIRNG 260
 QY 483 IYQMRADG 491

RESULT 8
 ID W22987 standard; Protein: 241 AA.

AC W22987;
 DT 08-OCT-1997 (first entry)
 DE Human serine protease 67 (SP67).
 KM Human; colon carcinoma; COLO 201; cell line; serine protease; SP67;
 KM screening; inhibitor; treatment; disease.
 OS Homo sapiens.
 PN J09149790-A.
 PD 10-JUN-1997.
 PF 24-JUL-1996; 212196.
 PR 29-SEP-1995; JP-275105.
 PR (SUNR) SUNTORY LTD.
 DR WPI: 97-357902/33.
 DR N-PSDB; T79128.
 PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
 PT study agent for treatment of various diseases
 PS Claim 1; Pages 12-13; 16pp; Japanese.
 CC The present sequence is the human colon carcinoma COLO 201
 CC cell line derived serine protease 67 (SP67), which can be used to
 CC screen for specific inhibitors, e.g. to search for, or study an
 CC agent for the treatment of various diseases.
 SQ Sequence 241 AA;

Query Match 17.7%; Score 649; DB 1; Length 241;

Best Local Similarity 40.1%; Pred. No. 3.95e-52;
 Matches 95; Conservative 56; Mismatches 76; Indels 10; Gaps 8;

DB 1 YVGGTADAGEPMQVSLALGCHICGASLSPMVLVAHCACTIDRGFRVSDPTOMTV 60
 QY 255 YVGGESALPGAMPQVSLH-VONVHVCSSITTPMIVTAHAC-VE-KL--NNPWHMTA 309
 DB 61 FLGLHDOSORSAAGVOERLLRIISHPEFNDPTFYDALLELEKPAEYSSNVRICLPD 120
 QY 310 FGLHDSQMFY-GA-GYQVKEVISHPNYDTSKKNNDIALMLKLPKPLTFNDLVKVCPLN 367
 DB 121 ASHVPFAGKAVTQWGHQYGTGALLIQKEIKVINTOTEN--LLPOQITPPMKCVG 178
 QY 368 PGMMLOPEQLCHISWGAT-EKRG-KTSEVLAARVLLIETORCSRYVDNLTPMICAG 427
 DB 179 FLSGGVDSQSGSGGPLSSVEADGRIFQGVVSWGDCGAKRRKPGVYTRLPFLRFMI 235
 QY 428 FLQGVNDSCQDSGSGPLVTSKNNIM-WLIGDTSWGSCKAKAYRPGVYGVNMTDM 483

PR 28-OCT-1997; US-063541.
PR 28-OCT-1997; US-063542.
PR 28-OCT-1997; US-063544.
PR 28-OCT-1997; US-063549.
PR 28-OCT-1997; US-063550.
PR 28-OCT-1997; US-063564.
PR 29-OCT-1997; US-063435.
PR 29-OCT-1997; US-063704.
PR 29-OCT-1997; US-063732.
PR 29-OCT-1997; US-063738.
PR 29-OCT-1997; US-063734.
PR 29-OCT-1997; US-064215.
PR 29-OCT-1997; US-063735.
PR 31-OCT-1997; US-063870.
PR 31-OCT-1997; US-064103.
PR 03-NOV-1997; US-064248.
PR 07-NOV-1997; US-064809.
PR 12-NOV-1997; US-065186.
PR 17-NOV-1997; US-065846.
PR 18-NOV-1997; US-065693.
PR 21-NOV-1997; US-066120.
PR 21-NOV-1997; US-066354.
PR 24-NOV-1997; US-066772.
PR 24-NOV-1997; US-066466.
PR 24-NOV-1997; US-066770.
PR 24-NOV-1997; US-066511.
PR 24-NOV-1997; US-066511.
PR 24-NOV-1997; US-066453.
PA (GEM) GENENTECH INC.
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
DR WPI: 99-229533/19.
DR N-PSDB; X53262.
PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
PS Claim 12: Fig 98, 320pp; English.
PS Y13444-03 represent secreted and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to growth
CC or survival of nerve cells including Parkinson's disease, Alzheimer's
CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 can be used in the treatment
CC of Usher Syndrome or Atrophica areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
SQ Sequence 317 AA;

Query Match 16.68; Score 608; DB 1; Length 317;
Best Local Similarity 37.48; Pred. No. 5,80e-48;
Matches 89; Conservative 54; Mismatches 88; Indels 7; Gaps 5;

DB 48 NRYVGGDSRDSSEMPVVSLOKNGTHHCAGSLTSRWVITAACFKNLKPKYLSVLG 107
OY 253 SRIYGGESALPGAMPVOSLHVONVHVGSSITPEVITVAACVEKPLNPNHMTAFAG 312
DB 108 AMOIGNPGSRKQVAVMEHPVYSKKEGACADIALVRLERSIQSESVLPICLPDASI 167
OY 313 ILNGSEMFY-GAGYQVEKVISHPYDSKTKN-NDIALMLKLOKPLTFNDLVKPVCLPBM 370
DB 168 HLPNTHCMISGMSIDQVPLPHPOTLOKLYKIIDSVCSHLYMGAGAGOPTDMLC 227
OY 371 MLOPEOLCWTISGMCATEKCK-T-SEVLNAAKVALLIETORCNSRY--VYDNLITPMAC 425

DB 228 AGTLEGEDRACLGDSGGPLMCGVNDGMLLAGIISGEGCAENRRGVYISLSAHSRY 285
OY 426 AGFLQGNVDSGCGDSGGLYTSKNNIMWLIGDTSWGSACAKYRPGVGNVFTDWT 483

RESULT 15
ID W64239 standard; Protein: 270 AA.
AC W64239;
DE 24-NOV-1998 (first entry)
DE gerbil homologue of mouse mCP-7 zymogen.
KW MCP-7 zymogen; gerbil; mast cell protease 7; trypsin-7;
KW blood clot; anticoagulant; myocardial infarction; reocclusion;
KW thromboembolism; cerebral embolism; thrombolysis; therapy.
OS Meriones sp.
FH Key
FT Location/Qualifiers
FT 25..270
FT /Label= Mat_protein

PN W09824886-A1.
PD 11-JUN-1998.
PE 25-NOV-1997; U21620.
PR 04-DEC-1996; US-032354.
PA (BGHM) BRIGHAM 6 WOMENS HOSPITAL.
PI Stevens RL;
DR WPI: 98-333308/29.
DR N-PSDB; V44326.

PT New compositions containing trypsin-7, e.g. mouse mast cell
PT protease-7 - are used to treat clot formation in e.g. myocardial
PT infarction, reocclusion following angioplasty or pulmonary
PT thromboembolism
PS Disclosure: Page 61, 92pp; English.
CC This is the amino acid sequence of the gerbil homologue of mouse
CC mast cell protease 7 (mCP-7, see also W64233), deduced from a
CC nucleic acid sequence (see V44326). Mature gerbil homologue of
CC mCP-7 can be used to prevent or inhibit fibrin clot formation in
CC vitro or in vivo. The invention provides: compositions comprising
CC an isolated trypsin-7 such as mCP-7, its homologues or humanised
CC mCP-7; a method for treating a blood clot by administering a
CC nucleic acid molecule that codes for a trypsin-7, or an expression
CC product, to decrease fibrinogen activity; a nucleic acid encoding a
CC serine protease (SP); and a method of producing a mature SP by
CC expressing the inactive zymogen in a host cell, and cleaving the
CC enterokinase susceptibility domain. The trypsin-7 polypeptides
CC can be used to treat disorders mediated by undesirable thrombus
CC clot formation such as myocardial infarction and reocclusion
CC following angioplasty of blood clots associated with pulmonary
CC thromboembolism, deep vein thrombosis, cerebral embolism, renal
CC vein and peripheral arterial thrombosis. They are also useful for
CC all surgical procedures that require decreased blood clots.
SQ Sequence 270 AA;

Query Match 16.58; Score 607; DB 1; Length 270;
Best Local Similarity 40.48; Pred. No. 7.33e-48;
Matches 99; Conservative 46; Mismatches 80; Indels 20; Gaps 15;

DB 26 IVGGQEA-PGNKPMPOVSLRANETVRRHFGGSLHPQVLTAAACVGPPIADP-N-KVR 82
OY 255 IVGGESALPGA-WPMQVSLHVQNV--HVCGSIIITPEVITVAACVEKPLNPNHMTAF 310
DB 83 VQ-LRKOYLTHHHLAVSRILTHPTFATONONADTALTELRKPVNISSVHVPSPAS 141
OY 311 AGILROSEMFY-GAGYQVEKVISHPYDSKTKNNDIALMLKLOKPLTFNDLVKPVCLPBM 369
DB 142 EFPFSGILCVTGMGNINDVSLPPEPLKEVQVAVVENOJCDLKHGVTGDNHIVR 201
OY 370 MLOPEOLCWTISGMCATEKCK-T-SEVLNAAKVALLIETORCNSRY--VYDNLITPMAC 421
DB 202 DMLICAGN-BGH-DSCGDSGGPLVCVNGTWTLOAGVNSGECALPNRPGITRYTYIL 259
OY 422 A-MICAGFLQGNVDSGCGDSGGLYTSKNNIMWLIGDTSWGSACAKYRPGVGNVFTDWT 480
DB 260 DWIHR 264
OY 481 DWIYR 485

[illegible]

QY	1366	ctgtatcacacacagccatgactctctgtccgcgcttcctctgcaggggaagctgactcttcgccag	1425
Db	379	CTAATCACACACAG-CAGTAACTGTGGCGGTTCTCTCAGGGGCTGTGCACCTTGGCCAG	321
QY	1426	gctgcacgttgaggggctctctgtcaactctgaagaaacaaatcatctgtgtgtctgaaggagat	1485
Db	320	GGAGACAACTGGAGGGCGCGCTGTACTTTGAACAATGGGATCTGGTGGCTGATTGGGAC	261
QY	1486	acaagcttgggtttctgtgcctgtgcacaaagcttacagaccagagctgtacggaatgtga	1545
Db	260	ACGAGCTGGGCTCGGGCTGTGCCAAGGCACTCAGACCTGGAATATACGGGAACGTGACG	201
QY	1546	gtattccagggctgtgatttatcgaaacaaatgagggcagacgcttaatccaatg-----	1600
Db	200	GTAATTTACAGATTGGAACTCTACGACAAATGAGGGCCAAAGCTAATCCAGCTGGCTTGT	141
QY	1600	-----tcttcgtctcttgacgttcgttttaacgaagaaacaaatgtgggtcgt-gtttgtc	1649
Db	140	CCCAAGACTCTCTTTGTGTCTTACACAACTTCTCTGAAAGAAACCAAGGGCTGAAATTTAAC	81
QY	1650	tccccgttcagatattcatctctagagaatgattccagagctacatcatcttattataacag	1709
Db	80	TTTCCTGGACAATATGTAACCTTTTGAGATGATTTGGAAAGGGCTTTCACCTTTATTAAAG	21
QY	1710	tga	1712
Db	20	TGA	18

Search completed: April 22, 2000, 05:45:44
Job time: 2823 sec

Query Match	24.4%	Score 895	DB 1	Length 798
Best Local Similarity	36.3%	Pred. No. 3,00e-71		
Matches 138	Conservative	84	Mismatches 141	Indels 17
				Gaps 16
Db	422	CKEDNFOCKD-GEICPLVNLCDGFPCHKDGSDAHCYRLFNGTSSGLVQFRIOST-WH	479	
Qy	113	CSNGIECDSSGTCINPNSMCDGSHCPGCGEDENRCVRLT-GP-NFI-LQVYSSORKSWM	169	
Db	480	VACAEKNTTIOISDDVQCLLGLTG-C-NSVYTFPS-TGGGPRVINTAP-NGLSLLT-PSQQ	535	
Qy	170	PVQODDMNENYGRAACRDMGYNKNFYSSGIVDSSGTSFMKLNTSAGNDVITKTHSDA	229	
Db	536	CLEDSLILQCNKSKCKLVTQEVSPKRIYGGSDSRGAPMVVALYFDDQVCGASLVS	595	
Qy	230	CSKSAVYSLFCT--ACGVNLNSSNQS-RIVGGBALPGAMPQVSLHYQVHVHCGGSIIT	286	
Db	596	RDMLVSAHACVYGRNME-PSKMKAVILGLHNASNLTSFOIETRLIDQIVINPHYKRRKN	654	
Qy	287	PEWIVTAACHCV-EKPLNNPMMHMTAFAGILRQSEFYGA-GYQ-VEKISHPNYDSKTKN	343	
Db	655	DIAHMLHEKVNNTDIQPICLPEBNVFPFPRGICISAGKALLYGOSTADVLOEADVPL	714	
Qy	344	DIALMKLOKELTFEDILKPVCLPFGPMLOPBDLCWISGAAEEKGKTSVLAARVLL	403	
Db	715	LSNEKCOGO-MPEXINENNYCAAGAEAGVDSGCGSGCLMCOENNRMLLAGTSPGYQ	773	
Qy	404	IEIQRCSKRYVDNLITPAHICAGFLOGVNDSCGDSGGPLVSKNNIMWILIDTISWGS	463	
Db	774	CALPNRPGVYARVPRTEWTI	793	
Qy	464	CANAYRPGVYGVNMFDTWI	483	
RESULT	2	STANDARD:	PRT:	798 AA.
ID	PCT-US94-00616-2			
XX	xxxxxx			
DT				
DE	Sequence 2, Application PC/TUS9400616			
XX				
CC	Sequence 2, Application PC/TUS9400616			
CC	GENERAL INFORMATION:			
CC	APPLICANT:			
CC	TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE			
CC	NUMBER OF SEQUENCES: 33			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: PCT/US94/00616			
CC	FILING DATE:			
CC	CLASSIFICATION:			
CC	INFORMATION FOR SEQ. ID NO: 2:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 798 amino acids			
CC	TYPE: amino acid			
CC	STRANDEDNESS: single			
CC	TOPOLOGY: linear			
CC	MOLECULE TYPE: protein			
SO	SEQUENCE 798 AA; 88937 MM; 3292434 CN;			
Db	Query Match	24.4%	Score 895	DB 3
	Best Local Similarity	36.3%	Pred. No. 3,00e-71	
	Matches 138	Conservative	84	Mismatches 141
				Indels 17
				Gaps 16
Db	422	CKEDNFOCKD-GEICPLVNLCDGFPCHKDGSDAHCYRLFNGTSSGLVQFRIOST-WH	479	
Qy	113	CSNGIECDSSGTCINPNSMCDGSHCPGCGEDENRCVRLT-GP-NFI-LQVYSSORKSWM	169	
Db	480	VACAEKNTTIOISDDVQCLLGLTG-C-NSVYTFPS-TGGGPRVINTAP-NGLSLLT-PSQQ	535	
Qy	170	PVQODDMNENYGRAACRDMGYNKNFYSSGIVDSSGTSFMKLNTSAGNDVITKTHSDA	229	
Db	536	CLEDSLILQCNKSKCKLVTQEVSPKRIYGGSDSRGAPMVVALYFDDQVCGASLVS	595	
Qy	230	CSKSAVYSLFCT--ACGVNLNSSNQS-RIVGGBALPGAMPQVSLHYQVHVHCGGSIIT	286	
Db	596	RDMLVSAHACVYGRNME-PSKMKAVILGLHNASNLTSFOIETRLIDQIVINPHYKRRKN	654	
Qy	287	PEWIVTAACHCV-EKPLNNPMMHMTAFAGILRQSEFYGA-GYQ-VEKISHPNYDSKTKN	343	
Db	655	DIAHMLHEKVNNTDIQPICLPEBNVFPFPRGICISAGKALLYGOSTADVLOEADVPL	714	
Qy	344	DIALMKLOKELTFEDILKPVCLPFGPMLOPBDLCWISGAAEEKGKTSVLAARVLL	403	
Db	715	LSNEKCOGO-MPEXINENNYCAAGAEAGVDSGCGSGCLMCOENNRMLLAGTSPGYQ	773	
Qy	404	IEIQRCSKRYVDNLITPAHICAGFLOGVNDSCGDSGGPLVSKNNIMWILIDTISWGS	463	
Db	774	CALPNRPGVYARVPRTEWTI	793	
Qy	464	CANAYRPGVYGVNMFDTWI	483	

```

Db      480 VACAEWTTQSDVOCCLLGATG-NSVSPTFS-TGGGPYVNLTAP-NSLSILT-PSOQ 535S
Oy      170 PYCQDMMNENTGRACRCDMGCKTKNNFYSOGCIYDDSGSTSPKMLTSSAGNDIYKKHSDA 229Y
Cc      CLEDSLILAQCNMYKSCGKLVTQEVSPRIYGGSDSRREGAMPVVALLYFDDOVCASLVS 599S
Db      536 CLEDSLILAQCNMYKSCGKLVTQEVSPRIYGGSDSRREGAMPVVALLYFDDOVCASLVS 599S
Oy      230 CSKRNAVYSLRCI--ACGVNLNSSHQ-S-RIVGGESALPGAMPQVSLHVQNVHVGCGSIIT 286Y
Cc      596 RDLVLVAACHCYGNMME-PSKWKAKVLGLHMASNLTSFOIEIRLDQVIYNPHYKRRKN 654S
Oy      287 PEWITTAHCV-EKPLANNPWHMTFAIGLRQSFEFYCA-GYO-VEKIISHNPDSKTNN 343Y
Db      655 DIAMNHLEMKNYNDIYDIPCLPEENGYEPPRGICSIAGMCALTYOGSTADVLOEAQVPL 714A
Oy      344 DIALMKLRQKLPFLENDLVKPCVLPBGMKLOEOELCMISGWCATEKEGTSEVLNAAYLL 409Y
Db      715 LSNKCOOO-APEXNITEBNVACAYEAGVDSCOGSDSGGPLMCOENNRMLLAGYTSFGYO 773S
Oy      404 IETQRCSRRIYYDNLITPAMLCACAFLOGINDSCOGSDSGGPLVTSKNNIMWLIDITSNGSG 466Y
Db      774 CALPNRPGEVYARVPTEMI 793
Oy      464 CAKAVRPGCVGNVMVFETDWI 483

RESULT          3           STANDARD;             PRT;            638 AA.
ID              US-08-681-151-3
XX              xxxxxx
XX              Sequence 3, Application US/08681151
Dt              Sequence 3, Application US/08681151
Dt              Patent No. 5868637
Cc              GENERAL INFORMATION:
Cc              APPLICANT: Au-Young, Janice
Cc              APPLICANT: Bandman, Olga
Cc              APPLICANT: Braxton, Scott Michael
Cc              APPLICANT: Goll, Surya
Cc              TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
Cc              NUMBER OF SEQUENCES: 4
Cc              CORRESPONDENCE ADDRESSES:
Cc              ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
Cc              STREET: 3174 Porter Drive
Cc              CITY: Palo Alto
Cc              STATE: CA
Cc              COUNTRY: US
Cc              ZIP: 94304
Cc              COMPUTER READABLE FORM:
Cc              MEDIUM TYPE: Diskette
Cc              COMPUTER: IBM Compatible
Cc              OPERATING SYSTEM: DOS
Cc              SOFTWARE: FASTSEQ Version 1.5
Cc              CURRENT APPLICATION DATA:
Cc              APPLICATION NUMBER: US/08/681,151
Cc              FILING DATE: Herewith
Cc              CLASSIFICATION: 435
Cc              PRIOR APPLICATION DATA:
Cc              APPLICATION NUMBER:
Cc              FILING DATE:
Cc              ATTORNEY/AGENT INFORMATION:
Cc              NAME: Billings, Lucy J.
Cc              REGISTRATION NUMBER: 36,749
Cc              REFERENCE/DOCKET NUMBER: PP-0074US
Cc              TELECOMMUNICATION INFORMATION:
Cc              TELEPHONE: 415-855-0555
Cc              TELEFAX: 415-845-4166
Cc              TELEX:
Cc              INFORMATION FOR SEQ ID NO: 3:
Cc              SEQUENCE CHARACTERISTICS:
Cc              LENGTH: 638 amino acids

```


Db 1141 aatgacattcgcgtatgaagctgcagaaagcctcctgactttcaacagactagtgaaacca 1200
Qy 1201 gtgtgtcgcgcacaccccgagatgtctgcagaccagaaacgctctgttggattccggg 1260
Db 1201 gtgtgtcgcgcacaccccgagatgtctgcagaccagaaacgctctgttggattccggg 1260
Qy 1261 tgggggagcagcagagaaagagagagaccctcaagagtgcttaagcctgcgcagagtgctt 1320
Db 1261 tgggggagcagcagagaaagagagagaccctcaagagtgcttaagcctgcgcagagtgctt 1320
Qy 1321 ctcatctgagacagagatgcaacagcagatagtctatgacaacctgatacacacagcc 1380
Db 1321 ctcatctgagacagagatgcaacagcagatagtctatgacaacctgatacacacagcc 1380
Qy 1381 atgattctgtccggtctctcgcagggagagctgcattcttcgcaaggtgcagagtggaagg 1440
Db 1381 atgattctgtccggtctctcgcagggagagctgcattcttcgcaaggtgcagagtggaagg 1440
Qy 1441 cctctgtcactctcgaagaaacatactcgtgtgcataagggatacaaaagctggggtctc 1500
Db 1441 cctctgtcactctcgaagaaacatactcgtgtgcataagggatacaaaagctggggtctc 1500
Qy 1501 ggcgtgcacaaagcttacagacagagtgtaaggaatgtgtatctcaaggaatgg 1560
Db 1501 ggcgtgcacaaagcttacagacagagtgtaaggaatgtgtatctcaaggaatgg 1560
Qy 1561 attatcgcacaaatgagggcagacggtcaatcacatggtctcgtctgagctgctgtt 1620
Db 1561 attatcgcacaaatgagggcagacggtcaatcacatggtctcgtctgagctgctgtt 1620
Qy 1621 tacaagaaacaaatgagggctgttctgtctcccgctgcatacttaactcttaagatgat 1680
Db 1621 tacaagaaacaaatgagggctgttctgtctcccgctgcatacttaactcttaagatgat 1680
Qy 1681 tcagagggcactctcttattataaacgtgtaactgtctgcacaaaaaaataaa 1738
Db 1681 tcagagggcactctcttattataaacgtgtaactgtctgcacaaaaaaataaa 1738

RESULT 2
US-09-526-996-1043
: Sequence 1043, Application US/09526996
: GENERAL INFORMATION:
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NOCLETIC ACID MOLECULES DERIVED FROM
: FILE REFERENCE: 1600.1088-001
: CURRENT APPLICATION NUMBER: US/09/526,996
: EARLIER FILING DATE: 2000-03-15
: EARLIER APPLICATION NUMBER: 60/125,469
: EARLIER FILING DATE: 1999-03-22
: EARLIER APPLICATION NUMBER: 60/125,560
: NUMBER OF SEQ ID NOS: 1296
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1043
: LENGTH: 3411
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-526-996-1043

Query Match 99.0%; Score 1721.4; DB 92; Length 3411;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggcgagagcagagagcagagagcagagagcagagagcagagagcagagagcagagagcagagat 60
Db 28 ggcgagagcagagagcagagagcagagagcagagagcagagagcagagagcagagagcagagat 87
Qy 61 attgaacattccagatacattacattactcagatgtctgtatgaacaagcaagatggctttg 120

Db 88 attgaacattccagatacattacattactcagatgtctgtatgaacaagcaagatggctttg 147
Qy 121 aactcaggggtcacacacacagcattatgtgaccttactatgaaaaacatagataacacaggaa 180
Db 148 aactcaggggtcacacacacagcattatgtgaccttactatgaaaaacatagataacacaggaa 207
Qy 161 aacctctaccgcaacagaccacgtgtgtcccaactgtctcagagtgatcatccggctcag 240
Db 208 aacctctaccgcaacagaccacgtgtgtcccaactgtctcagagtgatcatccggctcag 267
Qy 241 tactaccggtcccccgtgtcccccagtagcccccagaggtcccgcaagcttcccaacccc 300
Db 268 tactaccggtcccccgtgtcccccagtagcccccagaggtcccgcaagcttcccaacccc 327
Qy 301 gtctgtctgacagcagaccacaaatcccccacccgagagagtgtaacaccccaagataaaga 360
Db 328 gtctgtctgacagcagaccacaaatcccccacccgagagagtgtaacaccccaagataaaga 387
Qy 361 gcaactgtcatcaaccttgaaaccttgaggacctctcgtgtggagagctgtgcgtgcgtgc 420
Db 388 gcaactgtcatcaaccttgaaaccttgaggacctctcgtgtggagagctgtgcgtgcgtgc 447
Qy 421 ctactctggaagtctatgtggagcagaagtgctcaaacctcttgagataagatgacacccca 480
Db 448 ctactctggaagtctatgtggagcagaagtgctcaaacctcttgagataagatgacacccca 507
Qy 481 ggtacctgcatcaaacccctctaacctgtgtgataggcgtgtcacactgcgcgcggcgagg 540
Db 508 ggtacctgcatcaaacccctctaacctgtgtgataggcgtgtcacactgcgcgcggcgagg 567
Qy 541 gacagaaatcgggtgtgtctgcgcctctacaggaacaaactcaactcctcaaggctgaatcatc 600
Db 568 gacagaaatcgggtgtgtctgcgcctctacaggaacaaactcaactcctcaaggctgaatcatc 627
Qy 601 cagagaaagtctctgtgcacacctgtgtgcacagcagctgtgaaagaaactacagggcgagc 660
Db 628 cagagaaagtctctgtgcacacctgtgtgcacagcagctgtgaaagaaactacagggcgagc 687
Qy 661 gctctcagaggacattggtcctataagataattttctctcagaagagataatgtgatac 720
Db 688 gctctcagaggacattggtcctataagataattttctctcagaagagataatgtgatac 747
Qy 721 agcgatccacacagcttctatgaacatgaacacaaagtgcgcgacatgtgatacctataaa 780
Db 748 agcgatccacacagcttctatgaacatgaacacaaagtgcgcgacatgtgatacctataaa 807
Qy 781 aaacgtacacacagatgtgcgtgtctctcaaaagcagtggttctcttcgcgtgatacgc 840
Db 808 aaacgtacacacagatgtgcgtgtctctcaaaagcagtggttctcttcgcgtgatacgc 867
Qy 841 tgcggggtcgaacttgaaactcgaagccgcgcagagcagagattgtgagggagagcgcgctc 900
Db 868 tgcggggtcgaacttgaaactcgaagccgcgcagagcagagattgtgagggagagcgcgctc 927
Qy 901 ccgggggacctgtgcctgtgcaggtcagcctgtcagcagtgctcagaaagtgctgcggaagc 960
Db 928 ccgggggacctgtgcctgtgcaggtcagcctgtcagcagtgctcagaaagtgctgcggaagc 987
Qy 961 tccatcatcaaccccgagatgtgactgtgacacgcgcacacgtcgtgtgaaaaaacctctaac 1020
Db 988 tccatcatcaaccccgagatgtgactgtgacacgcgcacacgtcgtgtgaaaaaacctctaac 1047
Qy 1021 aatccatgtgcatgtgacagcattgtcgggagatttgagagaaacattctcatgtctcatga 1080
Db 1048 aatccatgtgcatgtgacagcattgtcgggagatttgagagaaacattctcatgtctcatga 1107
Qy 1081 ggcgagatacagaagtgaagaaagtattctcatccaatattagactccaagaccaaagaac 1140
Db 1108 ggcgagatacagaagtgaagaaagtattctcatccaatattagactccaagaccaaagaac 1167
Qy 1141 aatgaacattgcgtatgtgaagctgcagaaagcctcgtactttcaagacacttagtgaaacca 1200
Db 1168 aatgaacattgcgtatgtgaagctgcagaaagcctcgtactttcaagacacttagtgaaacca 1227

```

CC REFERENCE/DOCKET NUMBER: PF-0074US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 356 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC IMMEDIATE SOURCE:
CC LIBRARY: HEBAROT01
CC CLONE: 307474
CC
SQ SEQUENCE 356 AA; 39995 MW; 717455 CN;
SQ
Query Match 18.8%; Score 691; DB 2; Length 356;
Best Local Similarity 35.8%; Pred. No. 4,676-52;
Matches 118; Conservative 67; Mismatches 128; Indels 17; Gaps 16;
Db
25 HHVACADGMOEILLSOLACKOMGLGEPSTVKL-IOGEKEPRMLTSHSNMESLNGTTLHEL 83
169 HPCVODDWNENYGRACRDMGYKNMFYSSGCIYDDSGSTSPMKLNTS-AG-N-VDIYKKY 225
Db
84 VNGSGCESRKISLCTQDGCGRPAAMNKRILIGRTSRGRRPWCSDLOSEPSGHICG 143
226 -HSDGSSKAVAYSLSCI--ACGVNLNLSRSGRIYGGESALPGAMPWOVSLHVQ-NVHVC 281
Db
144 CVLLAKKKVLLVAHCEGREGRENAAYKKVVLGINLIDHPSEVFQTRF-VKTIILHRYSR 202
OY 282 GSIIPEPIVIAHCEVEPRPLNPNHMTAF-GLRQSMFYAGAGIYQEKVLSHPHYDSKT 340
Db
203 VDYDISIELSEDISETGYVRPCLLPNPEQWLEBDPYCYITGWSHMGNK-APFK-LOGE 260
OY 341 KNNIDALKLKQKPLFNLDLVKPVCLPNPGMMLDPEQLCWSGWGATECKGKTSVILNAK 400
Db
261 VRIISLEHGQS-Y-FDMKTTTTRMTCAGIESGYVDSCKGDSGGGLVCKPGCGRTTLELT 318
OY 401 VLLIETGHCNSRYYVD-NLITPMACAGFLOGNVDSCGDSGGGLVTSKNIMW-LIGDT 458
Db
319 SMGVCSPKSLVGPVYSNVSYFVEMIKROI 348
OY 459 SMGSGC-AKAYRPGVYGNVFTDMYIRQM 487
RESULT 6 STANDARD: PRT; 855 AA.
ID US-09-027-337-2
AC xxxxxx
XX
XX
XX
XX
DE Sequence 2, Application US/09027337B
XX
XX Sequence 2, Application US/09027337B
CC Patent No. 5972616
CC GENERAL INFORMATION:
CC APPLICANT: O'Brien, Timothy J.
CC APPLICANT: Tamimoto, Hirotooshi
CC TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed In
CC FILE REFERENCE: D6064
CC CURRENT APPLICATION NUMBER: US/09/027,337B
CC CURRENT FILING DATE: 1998-02-20
CC NUMBER OF SEQ ID NOS: 13
CC SEQ ID NO 2
CC LENGTH: 855
CC TYPE: PRT
CC ORGANISM: Homo sapiens
CC FEATURE:
CC OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
CC OTHER INFORMATION: 23 to 2589 of Sequence 1

```

[illegible]

```

OY 761 aaactgtacacagtgatgctgttcttccaaagcagtggttcttcttacgctgtatagcc 840
Db 816 AAACGTACACAGTATGATGCTGTTCTTCAAAAGAGTGCTGTTCTTACGCTATAGCC 875
OY 841 tgcggggtcgaacttgaaactcaagccgagcagagatgtgtggcgcgagagcgcgctc 900
Db 876 TGCGGGGTCAACTTGAACCAAGCCGCGAGACAGAGATGTGGCGGCGCAAGAGCGGCTC 935
OY 901 ccgggggctctgctccttgtaggttcagccgctgcagctccaaagctccagctgtagggc 960
Db 936 CCGGGGCTTGGCTGCGAGGTGACGCTGCACGTCACAAAGCTCCAGCTGTGCGGAGGC 995
OY 961 tccatcatcaccgcccgagtgatgctgtgacagccgcccacctgtgtgaaaaactctaac 1020
Db 996 TCCATCATCACCCTCCGAGTGCATGTCGACAGCCGCCACCTGCGTGAATAAACCTCTTAC 1055
OY 1021 aatccatgcatcttgagcagcattctgaggagattttgagacaactcttcatgtctatga 1080
Db 1056 AATCCATGGCATTTGAGACGCGATTTGGGGGATTTTGACAAATCTTTCATGTTCTATGGA 1115
OY 1081 gccggaatccaaagtcaaaaagtgtttctatccaaattatgacctccaaagccaaagac 1140
Db 1116 GCCGATACCAAGTGAATAAGTATTTCTCATCCAAATTAAGACTCCAAAGACCAAGAAC 1175
OY 1141 aatgacattgctgctgtagagctgtagaagcctctgaccttcaacgacctagtgaaaca 1200
Db 1176 AATGACATTGCCCTGATGAGCTGCGAAGACCTCTGACTTTACACAGACCTAGTGAACCA 1235
OY 1201 gtgtgtctgcccacacagcagcaatgtctgtagcagcagaaacagctctgtgattctcgg 1260
Db 1236 GGTGTGCTGCCACACCGACGATGATGTCGACGCCAGAACACCTGCTGAGATTTTCGGG 1295
OY 1261 tggggggccacagagagagagagagagagagagagagagagagagagagagagagag 1320
Db 1296 TGGGGGCGCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1355
OY 1321 ctcatctgagacacagagatgacacagacagatattgtctatgacaaactgttcaacagac 1380
Db 1356 CTCATTGACACACAGAGATGCAACAGACAGATATGCTATAGACAACTGATCAGACAGCC 1415
OY 1381 atgactgtgtccggtctctctgaggggaaagctgacttcttgcaggggtgacagtgaggg 1440
Db 1416 ATGACTGTGCGCGGCTTCTCTGACGGGAGCGTGAATTTTCCAGGGGTGACAGTGAAGG 1475
OY 1441 cctctgtgtaacttcgaaagaacatatctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
Db 1476 CCTGTGTCACCTTCGAAAGCAATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1535
OY 1501 ggcctgtgccaagaagcttaacagacagagagtgtagcggaaatgtgtgtgtgtgtgtgt 1560
Db 1536 GCGTGTGCCAAAGCTTACAGACAGAGAGTGTACGGGAATGTGTGTGTGTGTGTGTGTGT 1595
OY 1561 attatcgaacaaatgagggcagagcagacacacatgtgtctgtccttgaacgtctt 1620
Db 1596 ATTATTCGAACAATAGGCGCAAGCGCTAATCCATGCTGTCTTGTGTGTGTGTGTGTGT 1655
OY 1621 taagaagaaacaaatgggctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
Db 1656 TACAAGAAACAATGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1715
OY 1681 taagaagtaactcttattataaacagtgaaactgtgtgtgtgtgtgtgtgtgtgtgt 1723
Db 1716 TCAGAGGTCACTTCAATTTTATTAACAGTGAACCTGTGTGTGTGTGTGTGTGTGTGT 1758

```

```

RESULT 4
PCT-US99-24331-8
; Sequence 8, Application PC/TUS9924331
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Recipon, Herve
; APPLICANT: Caferkey, Robert

```

```

; APPLICANT: diadexus, LLC
; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
; FILE REFERENCE: DEX-0052
; CURRENT APPLICATION NUMBER: PCT/US99/24331
; CURRENT FILING DATE: 1999-10-19
; EARLIER APPLICATION NUMBER: 60/104,737
; EARLIER FILING DATE: 1998-10-19
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 3443
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US99-24331-8

```

```

Query Match 98.4%; Score 1710.4; DB 1; Length 3443;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

```

OY 1 ggcggaggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 60
Db 38 ggcggaggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 97
OY 61 attgaacattccagatatacattactactcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
Db 98 attgaacattccagatatacattactactcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 157
OY 121 aactcaggggtccacacagcatttggacctactatgaataacatgtgatccaacaggaa 180
Db 158 aactcaggggtccacacagcatttggacctactatgaataacatgtgatccaacaggaa 217
OY 181 aaccctatccggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
Db 218 aaccctatccggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 277
OY 241 tactaccggtcccggtgtgcccagtaagcccgagggttctgtacgacaggtcttcaacccc 300
Db 278 tactaccggtcccggtgtgcccagtaagcccgagggttctgtacgacaggtcttcaacccc 337
OY 301 gtgcgttcgacgacgacccaataccccaatcccgagacagtggtgacactcacaagaactaaga 360
Db 338 gtgcgttcgacgacgacccaataccccaatcccgagacagtggtgacactcacaagaactaaga 397
OY 361 gcaactgtgatacacttggaccttgcctgcctgcctgcctgcctgcctgcctgcctgcctgc 420
Db 398 gcaactgtgatacacttggaccttgcctgcctgcctgcctgcctgcctgcctgcctgcctgc 457
OY 421 ctactctggaagtctatggagcagcaagtgctccaaactctgagatagagtgcaactctca 480
Db 458 ctactctggaagtctatggagcagcaagtgctccaaactctgagatagagtgcaactctca 517
OY 481 ggtacctgtacaaaccccttaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
Db 518 ggtacctgtacaaaccccttaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 577
OY 541 gacagaaatcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 578 gacagaaatcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 637
OY 601 cagaagaagctccgagacccctgtgtgccaagacagatggaagaaactacagggcgggcg 660
Db 638 cagaagaagctccgagacccctgtgtgccaagacagatggaagaaactacagggcgggcg 697
OY 661 gctctgagggagatggtgtataaagaaattttactactagccaagaaatagtgatgac 720
Db 698 gctctgagggagatggtgtataaagaaattttactactagccaagaaatagtgatgac 757
OY 721 agcggatccacagccttatgtgaactgaaacacaaagtgccgggcaatgtcgatatctataa 780
Db 758 agcggatccacagccttatgtgaactgaaacacaaagtgccgggcaatgtcgatatctataa 817

```

OY 263 PGMAMPQVSLHVNQVCGSITTPETIYTAHCEKPLNPNHMTAFACILROSEMFY 322
DB 229 VJGVOAVIYHGGYLPFRPTIDENSNDLALVHSSLPLETEYTOPVCLPAAQCALVDK 288
OY 323 AGVOEKVISHPRV-----DSKTK-N-NDIALMLQKPLTFNDLVKVCJLPNPGMMLQPEQ 376
DB 289 VCTVTGNTQFYGOQAMVLOEARVPILISNEVCNSPDEFGNQIKPKMFCAGYBEGGIDAC 348
OY 377 LCMWSMGATEEKGKTSVYNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNVDS 436
DB 349 QGDSGPPFCEDSISGTSRRLGCIYVWGCALARKPGVYTVTDEREMIFKAITHS 407
OY 437 QGDSGPPV--TSKNNI--WMLIGTWSGSCAKATRPVGYGNVFTDWTYROMRAD 491

RESULT 10
ID US-08-326-785-1 STANDARD: PRT: 812 AA.
AC xxxxxx
DT
DT
DT
DE Sequence 1, Application US/08326785
CC
CC Sequence 1, Application US/08326785
CC Patent No. 5792845
CC GENERAL INFORMATION:
CC APPLICANT: Polkman, Judah
CC APPLICANT: O'Reilly, Michael
CC TITLE OF INVENTION: Angiostatin and Method of Use
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Jones & Askew
CC STREET: 191 Peachtree Street, 37th Floor
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: USA
CC ZIP: 30303-1769
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50
CC COMPUTER: Macintosh
CC OPERATING SYSTEM: 7.0
CC SOFTWARE: Microsoft Word
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/326,785
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/248,629
CC FILING DATE: 04/26/94
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Larry W. Stults, Ph.D.
CC REGISTRATION NUMBER: 34,025
CC REFERENCE/DOCKET NUMBER: 05213-0121
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-818-3700
CC TELEFAX: 404-818-3799
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 812
CC TYPE: amino acid
CC TOPOLOGY: linear
CC SEQUENCE 812 AA; 90846 MW; 3411937 CN;

Query Match 17.38; Score 634; DB 1; Length 812;
Best Local Similarity 41.48; Pred. NO. 9.85e-47;
Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;

DB 553 DYCDIPLCASAFSECGKPOVEPKKCGRVYGCVANPHSPMPQISLRTFTGQHFCCGT 612
OY 228 DACSSRAVYSLRCLACG-VALNNSR-QSRIVGGSALPGAMPQVSLHVG-NV-HVCGGS 283
DB 613 LJAPENVLTAHCLXS-SRP-EF--YKVIL-GAHEEYINGLDVQE-ISAVALILEPNR 666

OY 284 ITTPENITVAHCEKPLNPNHMTAFACILROSEMFYAGIOVEKVISHPNIDSKTKNN 343
DB 667 DIALKLRSRATITDVIDPACLPSPVYVADRTICYITGMEFTGFGAGR-LKEAQLPV 725
OY 344 DIAMKLQKPLTFNDLVKVCJLPNPGMMLQPEQLCMWSMGATEEKGKTSVYNAKAVLL 403
DB 726 IENKVCN-RVEYLNVRKSTELCAGQLAGVDSQGDGAPLVCFEKDKYILQGVTSNGL 784
OY 404 IETORCNSRYV-DNLITPAMICAGFLQGNVDSQGDGAPLVTSKNINIMLIGDTSMS 462
DB 785 GCARPKNPGYVAVSRFVDMIEREMNN 812
OY 463 GCARAYRPGVYGNVFTDWTYROMRAD 490

RESULT 11
ID US-08-248-629A-1 STANDARD: PRT: 812 AA.
AC xxxxxx
DT
DT
DT
DE Sequence 1, Application US/08248629A
CC
CC Sequence 1, Application US/08248629A
CC Patent No. 5639725
CC GENERAL INFORMATION:
CC APPLICANT: Polkman, Judah
CC APPLICANT: O'Reilly, Michael
CC TITLE OF INVENTION: Angiostatin and Method of Use
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Jones & Askew
CC STREET: 191 Peachtree Street, 37th Floor
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: USA
CC ZIP: 30303-1769
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50
CC COMPUTER: Macintosh
CC OPERATING SYSTEM: 7.0
CC SOFTWARE: Microsoft Word
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/248,629A
CC FILING DATE: 04/26/94
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Larry W. Stults, Ph.D.
CC REGISTRATION NUMBER: 34,025
CC REFERENCE/DOCKET NUMBER: 05213-0120
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-818-3700
CC TELEFAX: 404-818-3799
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 812
CC TYPE: amino acid
CC TOPOLOGY: linear
CC SEQUENCE 812 AA; 90846 MW; 3411937 CN;

Query Match 17.38; Score 634; DB 1; Length 812;
Best Local Similarity 41.48; Pred. NO. 9.85e-47;
Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;

DB 553 DYCDIPLCASAFSECGKPOVEPKKCGRVYGCVANPHSPMPQISLRTFTGQHFCCGT 612
OY 228 DACSSRAVYSLRCLACG-VALNNSR-QSRIVGGSALPGAMPQVSLHVG-NV-HVCGGS 283

|||||
Db 818 aaactgaccacagatgacgtctgtcttccaagaagctgttcttctaagcgtataagcc 877
Qy 841 tgcgggggccaacttgaactcaagccgcagagacagatgtgtggcgagagcgcgcc 900
Db 878 tgcgggggccaacttgaactcaagccgcagagacagatgtgtggcgagagcgcgcc 937
Qy 901 ccgggggagcgtgacct-gcgaggtacgcgtgcacgtccagaagcgtccagctgtcggaag 959
Db 938 ccgggggagcgtgacctgtggcgaggtacgcgtgcacgtccagaagcgttcgaagtg 997
Qy 960 ctccatcaaccaccccgagtgagatcgtagacagccgcacgtcggtggaagaaaccttca 1019
Db 998 ctccatcaaccaccccgagtgagatcgtagacagccgcacgtcggtggaagaaaccttca 1057
Qy 1020 caatccatgagatgtgagagcattgtcggggagttttgagaaacttttattttatgt 1079
Db 1058 caatccatgagatgtgagagcattgtcggggagttttgagaaacttttattttatgt 1117
Qy 1080 agccggaatacgaagtagaagaaagtgattctcatccaataatgactccaagaacgaaga 1139
Db 1118 agccggaatacgaagtagaagaaagtgattctcatccaataatgactccaagaacgaaga 1177
Qy 1140 caatgacaatgcgtctgagagagtgtagaagcctctgacttccaagacatagtgaaac 1199
Db 1178 caatgacaatgcgtctgagagagtgtagaagcctctgacttccaagacatagtgaaac 1237
Qy 1200 agtggtctgcccacacccagacatgtgtctgcagccagaaacgctctgtgattccgg 1259
Db 1238 agtggtctgcccacacccagacatgtgtctgcagccagaaacgctctgtgattccgg 1297
Qy 1260 gtcgggggccaacggagagagagagagcctcagaaggtctggaagcgtcgcaaggtgt 1319
Db 1298 gtcgggggccaacggagagagagagagcctcagaaggtctggaagcgtcgcaaggtgt 1357
Qy 1320 tctcctctgagaaacagagaaatgcgaacagagatacttctatgacaacgtgataccagc 1379
Db 1358 tctcctctgagaaacagagaaatgcgaacagagatacttctatgacaacgtgataccagc 1417
Qy 1380 catgactgtgcggcgtctctgcagaggggaagctgcattcttcagaggtgagacgtgagag 1439
Db 1418 catgactgtgcggcgtctctgcagaggggaagctgcattcttcagaggtgagagag 1477
Qy 1440 ggcctctgctcactctgaaagaaataatactgtggtcgtgagagggatataagcttgagttc 1499
Db 1478 ggcctctgctcactctgaaagaaataatactgtggtcgtgagagggatataagcttgagttc 1537
Qy 1500 tgcgtctgccaagagcttcaagacagagagtgtagaagtgtagatgtagatgtagatgtagatg 1559
Db 1538 tgcgtctgccaagagcttcaagacagagagtgtagaagtgtagatgtagatgtagatgtagatg 1597
Qy 1560 gattatcgaacaaatgagggcgagacggtctaaatccacatgcttctgcttgaagcgtct 1619
Db 1598 gattatcgaacaaatgagggcgagacggtctaaatccacatgcttctgcttgaagcgtct 1657
Qy 1620 ttacagaagaacaaatgagggcgagtttcttcccggtgtagatgtagatgtagatgtagatg 1679
Db 1658 ttacagaagaacaaatgagggcgagtttcttcccggtgtagatgtagatgtagatgtagatg 1717
Qy 1680 ttacagaagaacaaatgagggcgagtttcttcccggtgtagatgtagatgtagatgtagatg 1723
Db 1718 ttacagaagaacaaatgagggcgagtttcttcccggtgtagatgtagatgtagatgtagatg 1761

RESULT 6
PCT-US99-14622-29

; Sequence 29, Application PC/TUS9914622

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavligian, Sean V.

; APPLICANT: Teng, David H.-F.

; APPLICANT: Myriad Genetics, Inc.

; TITLE OF INVENTION: TMRPS2 is a Tumor Suppressor

; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: PCT/US99/14622
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US99-14622-29

Query Match 95.2%; Score 1655.2; DB 1; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 56 gtcataatgaaacatccagatccatatactatctcgtatgtgtgataacagcaagatgg 115
Db 1 gtcataatgaaacatccagatccatatactatctcgtatgtgtgataacagcaagatgg 60
Qy 116 ctttgaactcaggttccacacagctatttgagacttactatgaaacacatggtatccaac 175
Db 61 ctttgaactcaggttccacacagctatttgagacttactatgaaacacatggtatccaac 120
Qy 176 cggaaaaaccctatccgcagacagccacgtgtgtccacgttctaaaggtgacatccg 235
Db 121 cggaaaaaccctatccgcagacagccacgtgtgtccacgttctaaaggtgacatccg 180
Qy 236 ctcaagtaactaccgtcccggtcccgagtaagcccgaggggttcctgagcagagcttcca 295
Db 181 ctcaagtaactaccgtcccggtcccgagtaagcccgaggggttcctgagcagagcttcca 240
Qy 296 accccgtctgtcagacagcccaatcccaatcccgagagtggtgacacccaagaacta 355
Db 241 accccgtctgtcagacagcccaatcccaatcccgagagtggtgacacccaagaacta 300
Qy 356 agaaagcaatgtgcmataccttggaaccttgaggaccttccgttgaggagctgtgcgcg 415
Db 301 agaaagcaatgtgcmataccttggaaccttgaggaccttccgttgaggagctgtgcgcg 360
Qy 416 ctgagcctactctggaagttcattgagcagcaagtgctccaaaccttggaagtagatgcagt 475
Db 361 ctgagcctactctggaagttcattgagcagcaagtgctccaaaccttggaagtagatgcagt 420
Qy 476 cctcaggtactcgtcatcaaaccccttaactgtgtgtgagtggtgtcaacatgcgccgag 535
Db 421 cctcaggtactcgtcatcaaaccccttaactgtgtgtgagtggtgtgtcaacatgcgccgag 480
Qy 536 gggagagacgaagatcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 595
Db 481 gggagagacgaagatcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
Qy 596 catctcagaagagatccctctgacccctgtgtgccaagacagctggaagcaagatccagc 655
Db 541 catctcagaagagatccctctgacccctgtgtgccaagacagctggaagcaagatccagc 600
Qy 656 gggcggtcgtcagggagacatggtgctataagaaataattttactctagccaagaaatggt 715
Db 601 gggcggtcgtcagggagacatggtgctataagaaataattttactctagccaagaaatggt 660
Qy 716 atgacagcgagatccacagcttatagaactataaacaagaagtcgggcaatgtcgatatct 775
Db 661 atgacagcgagatccacagcttatagaactataaacaagaagtcgggcaatgtcgatatct 720
Qy 776 ataaaaaactgtgacacagatgtagctgtcttccaagaacagtggtttcttaacgtcta 835
Db 721 ataaaaaactgtgacacagatgtagctgtcttccaagaacagtggtttcttaacgtcta 780
Qy 836 tagcgtcgggggttcaacttgaactcaagccgcagagcagagatgtgtggcgagagcg 895
Db 781 tagcgtcgggggttcaacttgaacttcaagccgcagagcagagatgtgtggcgagagcg 840


```

Oy 956 gaggtccatcacacccccggatgctgacagccgccccactgctggaanaacctc 1015
    |||||||
Db 901 gaggtccatcacacccccggatgctgacagccgccccactgctggaanaacctc 960
Oy 1016 ttaacaatcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 1075
    |||||||
Db 961 ttaacaatcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 1020
Oy 1076 atgagcccgagataccagaatgagaagaatgattctcattccaaatattatgattcagaacca 1135
    |||||||
Db 1021 atgagcccgagataccagaatgagaagaatgattctcattccaaatattatgattcagaacca 1080
Oy 1136 agaaacaatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 1195
    |||||||
Db 1081 agaaacaatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 1140
Oy 1196 aaccagtgctgctgcccacacccagcatgcatgctgcaagcaagaactgctgcatgtt 1255
    |||||||
Db 1141 aaccagtgctgctgcccacacccagcatgcatgctgcaagcaagaactgctgcatgtt 1200
Oy 1256 ccgggtgaggggccacccagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 1315
    |||||||
Db 1201 ccgggtgaggggccacccagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 1260
Oy 1316 tggcttcattgagacagacagagatgcaacagagatgctatgtaacacccatgacacac 1375
    |||||||
Db 1261 tggcttcattgagacagacagagatgcaacagagatgctatgtaacacccatgacacac 1320
Oy- 1376 cagccatgacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1435
    |||||||
Db 1321 cagccatgacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1380
Oy 1436 gagggctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1495
    |||||||
Db 1381 gagggctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1440
Oy 1496 gttccggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1555
    |||||||
Db 1441 gttccggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1500
Oy 1556 actgattatcgacaacatgagggagagagagagagagagagagagagagagagagagagagag 1615
    |||||||
Db 1501 actgattatcgacaacatgagggagagagagagagagagagagagagagagagagagagagag 1560
Oy 1616 cgtttacaagaagaacaaatgagggctgctgctgctgctgctgctgctgctgctgctgctgct 1675
    |||||||
Db 1561 cgtttacaagaagaacaaatgagggctgctgctgctgctgctgctgctgctgctgctgctgct 1620
Oy 1676 atgattcagaggtcattcatttattatbaaagtgaaactgctgctgctgctgctgctgctgctgct 1723
    |||||||
Db 1621 atgattcagaggtcattcatttattatbaaagtgaaactgctgctgctgctgctgctgctgctgct 1668

```

RESULT 8

```

US-09-323-597-3
; Sequence 3, Application US/09323597
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel E
; APPLICANT: Hubert, Rene S
; APPLICANT: Leong, Kahan
; APPLICANT: Raitano, Arthur B
; APPLICANT: Saifan, Douglas C
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS AND THERAPY OF
; FILE REFERENCE: 1703-007, US1
; CURRENT APPLICATION NUMBER: US/09/323,597
; EARLIER FILING DATE: 1998-06-01
; EARLIER APPLICATION NUMBER: 60/087,598
; EARLIER FILING DATE: 1998-06-01
; EARLIER APPLICATION NUMBER: 60/091,474
; EARLIER FILING DATE: 1998-06-29
; EARLIER APPLICATION NUMBER: 60/129,521

```

```

; EARLIER FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATON INFORMATION:
; JOURNAL: Genomics
; VOLUME: 44
; PAGES: 309-320
; DATE: 1997
US-09-323-597-3

```

```

Query Match          95.2%  Score 1655.2  DB 48:  Length 2479;
Best Local Similarity 99.5%  Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

Oy 56 gtcaattgaacattccagataccatcatcattactcgtatgctgctgatacagaagaatg 115
    |||||||
Db 1 gtcaattgaacattccagataccatcatcattactcgtatgctgctgatacagaagaatg 60
Oy 116 ctgtgaactcaggggtcacacacagcatattggaaccttactatgaaacacatgataccaac 175
    |||||||
Db 61 ctgtgaactcaggggtcacacacagcatattggaaccttactatgaaacacatgataccaac 120
Oy 176 cggaaaaacccctatcccgacacagccactgctgctcccaactgctcagaggtgcatccgg 235
    |||||||
Db 121 cggaaaaacccctatcccgacacagccactgctgctcccaactgctcagaggtgcatccgg 180
Oy 236 ctcaactactaccgctcccgctgctcccgagtagcccgaggggtccctgacagagcttcca 295
    |||||||
Db 181 ctcaactactaccgctcccgctgctcccgagtagcccgaggggtccctgacagagcttcca 240
Oy 296 accccgtcgtctgacagcagacccaaatcccaatccggagcaggtgtgcaacctcaaaagacta 355
    |||||||
Db 241 accccgtcgtctgacagcagacccaaatcccaatccggagcaggtgtgcaacctcaaaagacta 300
Oy 356 agaaagacactgtgcatcaccctctgacccctggggagacctctctgtggagactgctgctgctg 415
    |||||||
Db 301 agaaagacactgtgcatcaccctctgacccctggggagacctctctgtggagactgctgctgctg 360
Oy 416 ctggcctactctggaagtcatagggcagcaagtgctccaaactctggaataagtgagtgagact 475
    |||||||
Db 361 ctggcctactctggaagtcatagggcagcaagtgctccaaactctggaataagtgagtgagact 420
Oy 476 cctcagtaactgcatcacaaccccttaactggtgtgtagtgagtgtagcaactgcccggcg 535
    |||||||
Db 421 cctcagtaactgcatcacaaccccttaactggtgtgtagtgagtgtagcaactgcccggcg 480
Oy 536 gggaggaagagaatcgtgtgtgtgtgctcctcctcagggccaaactctcctcaagtgact 595
    |||||||
Db 481 gggaggaagagaatcgtgtgtgtgtgctcctcctcagggccaaactctcctcaagtgact 540
Oy 596 catctcagaagagatcctgggacccctgtgtgtgccaagagcactggaacagagaactcagggc 655
    |||||||
Db 541 catctcagaagagatcctgggacccctgtgtgtgccaagagcactggaacagagaactcagggc 600
Oy 656 gggcggtcctgcaaggagacatgagcctaagaatlaatttactctcagcaagaagatagtg 715
    |||||||
Db 601 gggcggtcctgcaaggagacatgagcctaagaatlaatttactctcagcaagaagatagtg 660
Oy 716 atgacagcggatccacagcttatgaaactgaaacagaagtgccggacatgcatatct 775
    |||||||
Db 661 atgacagcggatccacagcttatgaaactgaaacagaagtgccggacatgcatatct 720
Oy 776 ataaaaaactgacacagatgtagctgtcttctcaaaagcagtggttctcttaacgtgta 835
    |||||||
Db 721 ataaaaaactgacacagatgtagctgtcttctcaaaagcagtggttctcttaacgtgta 780
Oy 836 tagcctgagggtcgaactgaaactcaaaccccgcaagagagagatgtggcgcgagagcg 895
    |||||||

```


[illegible]

```

? RESULT 10
? PCT-US99-14622-1
? Sequence 1 Application PC/TUS9914622
? GENERAL INFORMATION:
? APPLICANT: Wong, Alexander K.C.
? APPLICANT: TAVLIGIAN, Sean V.
? APPLICANT: Teng, David H.-F.
? TITLE OF INVENTION: TMPRSS2, Inc.
? FILE REFERENCE: 2318-202
? CURRENT APPLICATION NUMBER: PCT/US99/14622
? CURRENT FILING DATE: 1999-07-01
? EARLIER APPLICATION NUMBER: US 60/091,044
? EARLIER FILING DATE: 1998-06-29
? NUMBER OF SEQ ID NOS: 33
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 1479
? TYPE: DNA
? ORGANISM: Homo sapiens

```

```

FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
FEATURE:
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession NO. U75329.
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
PCT-US99-14622-1

```

Query Match	Similarity	85.0%	Score	1477.4	DB 1	Length	1479
Best Local	Similarity	99.9%	Pred.	NM. 0'			
Matches	1478	Conservative	0	Mismatches	1	Indels	0
							Gaps
							0
QY	112	atgctcttgaactcaggtcaggtcacacacaggtatgtgacctactatgatgaataacatgatatac	171				
DB	1	atgctcttgaactcaggtcaggtcacacacaggtatgtgacctactatgatgaataacatgatatac	60				
QY	172	caaacggaaaaacccctatccgcgacagccactgtgtgtccaccactctctacagagtgat	231				
DB	61	caacggaaaaacccctatccgcgacagccactgtgtgtccaccactctctacagagtgat	120				
QY	232	ccggtcagttactaacccgtgcccccggtgcacacagttacgcccagaggttctctgcgcagagt	291				
DB	121	ccggtcagttactaacccgtgcccccggtgcacacagttacgcccagaggttctctgcgcagagt	180				
QY	292	tccaaaccccgctgtctgtcgcgcagagcccaaatcccatccgggacaggtgtgcacacttaag	351				
DB	181	tccaaaccccgctgtctgtcgcgcagagcccaaatcccatccgggacaggtgtgcacacttaag	240				

DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SERINE PROTEASE HEPsin (EC 3.4.21.-).
 GN HPN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN
 RP
 RC
 RA MEDLINE; 93305733.
 RA FARLEY D., REYMOND F., NICK H.;
 RT "Cloning and sequence analysis of rat hepsin, a cell surface serine
 RT proteinase."
 RT Blochim. Biophys. Acta 1173:350-352(1993).
 CC
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 CC OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X70900; CAA50256.1; -
 CC PIR; S32013; S32013.
 CC PIR; S33777; S33777.
 CC HSSP; P00763; 1PMK.
 CC PROSITE; PS00134; TRYPsin_HIS; 1.
 CC PROSITE; PS00135; TRYPsin_SER; 1.
 CC PFAM; PF00089; trypsin; 1.
 CC Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 CC CHAIN 1 161
 CC DOMAIN 1 16
 CC TRANSMEM 17 43
 CC
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC CATALYTIC CHAIN (POTENTIAL).
 CC
 CC DOMAIN 44 416
 CC ACT_SITE 162 416
 CC ACT_SITE 202 202
 CC ACT_SITE 236 236
 CC ACT_SITE 352 352
 CC DISULFID 152 276
 CC DISULFID 187 203
 CC DISULFID 321 337
 CC DISULFID 348 380
 CC CARBOHYD 111 111
 CC
 CC SEQUENCE 416 AA; 44926 MW; 2F8A69FB CRC32;
 SO
 Query Match 17.83; Score 652; DB 1; Length 416;
 Best Local Similarity 40.68; Pred. No. 1.27e-142;
 Matches 101; Conservative 52; Mismatches 83; Indels 13; Gaps 9;

OY 427 GFLQGNVDSQGDGSGPLY-TSK-NNI--WMLIGDTWSGSGCAKAYRPGVGNVFTDW 482
 DB 399 IFOAIRKHS 407
 OY 483 IYQMRADG 491
 RESULT 14
 ID PLMN_MOUSE STANDARD; PRT; 812 AA.
 AC P20918;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PLASMINOGEN PRECURSOR (EC 3.4.21.7) (CONTAINS: ANGIOSTATIN).
 GN PLG.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN
 RP
 RA MEDLINE; 91184812.
 RA DEGEN S.J., BELL S.M., SCHAEFER L.A., ELLIOTT R.M.;
 RT "Characterization of the cDNA coding for mouse plasminogen and
 RT localization of the gene to mouse chromosome 17.";
 RT Genomics 8:49-61(1990).
 RN
 RP
 RA CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
 RA MEDLINE; 95042728.
 RA O'REILLY M.S., HOLMGREN L., SHING Y., CHEN C., ROSENTHAL R.A.,
 RA MOSS M., LANE W.S., CAO Y., SAGE E.H., FOLKMAN J.;
 RT "Angiostatin: a novel angiogenesis inhibitor that mediates the
 RT suppression of metastases by a Lewis lung carcinoma."
 RT Cell 79:315-328(1994).
 CC
 CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
 CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
 CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
 CC AND INFLAMMATION. IN OVULATION IT WEAKENS THE WALLS OF THE
 CC GRAFVIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
 CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
 CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
 CC LAMININ AND VON WILLEBRAND FACTOR.
 CC
 CC -1- FUNCTION: ANGIOSTATIN IS AN ANGIOGENESIS INHIBITOR THAT BLOCKS
 CC NEOVASCULARIZATION AND GROWTH OF EXPERIMENTAL PRIMARY AND
 CC METASTATIC TUMORS IN VIVO.
 CC
 CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASINOGEN
 CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
 CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
 CC
 CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
 CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
 CC
 CC -1- MISCELLANEOUS: IN THE PRESENCE OF THE INHIBITOR, THE ACTIVATION
 CC INVOLVES ONLY CLEAVAGE AFTER ARG-581, RESULTING IN 2 CHAINS HELD
 CC TOGETHER BY 2 DISULFIDE BONDS. WITHOUT THE INHIBITOR, THE
 CC ACTIVATION INVOLVES ALSO REMOVAL OF THE ACTIVATION PEPTIDE.
 CC
 CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
 CC
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J04766; AAA50168.1; -
 CC PIR; A38514; A38514.
 CC HSSP; P00747; 1PMK.
 CC MGD; MGI:97620; PLG.
 CC PROSITE; PS00021; KRINGLE_1; 4.
 CC PROSITE; PS00134; TRYPsin_HIS; 1.
 CC PROSITE; PS00135; TRYPsin_SER; 1.
 CC PROSITE; PS50070; KRINGLE_2; 5.

```

FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
US-09-342-749-1

```

```

Query Match      85.0%; Score 1477.4; DB 49; Length 1479;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 112 atgcttgaactcaagctcaccacagctatctgacacttatgaaacacatgatac 171
DB 1 atgcttgaactcaagctcaccacagctatctgacacttatgaaacacatgatac 60
QY 172 caacggaaacccctatcccgacagaccactgtgtcccaactgtctacagagtgcac 231
DB 61 caacggaaacccctatcccgacagaccactgtgtcccaactgtctacagagtgcac 120
QY 232 cgggctcaatctaccgctcccgctcccgctcccgctcccgctcccgctcccgctcc 291
DB 121 cgggctcaatctaccgctcccgctcccgctcccgctcccgctcccgctcccgctcc 180
QY 292 tccaaaccccgctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 351
DB 181 tccaaaccccgctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
QY 352 actaagaagaacacgtgtgcacacacacacacacacacacacacacacacacacac 411
DB 241 actaagaagaacacgtgtgcacacacacacacacacacacacacacacacacacac 300
QY 412 ggcgcgtgcctctctcgtgaagtctcagggcagcagtgctccaaactctggatagatgc 471
DB 301 ggcgcgtgcctctctcgtgaagtctcagggcagcagtgctccaaactctggatagatgc 360
QY 472 gactcctcagctacacgtgcacacacacacacacacacacacacacacacacacac 531
DB 361 gactcctcagctacacgtgcacacacacacacacacacacacacacacacacacac 420
QY 532 ggcggggagagacgagaaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 591
DB 421 ggcggggagagacgagaaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 592 tactcatctcagagagagatctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 651
DB 481 tactcatctcagagagagatctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
QY 652 gggcgggcgcgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 711
DB 541 gggcgggcgcgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
QY 712 gtgagctgacagcggatccacacagcttattgaaactgaaacaaagtgcggcaatgtcagat 771
DB 601 gtgagctgacagcggatccacacagcttattgaaactgaaacaaagtgcggcaatgtcagat 660
QY 772 attctataaaactgtacacacagcttattgtcttcttcaaaagcaatgttcttcaagc 831
DB 661 attctataaaactgtacacacagcttattgtcttcttcaaaagcaatgttcttcaagc 720
QY 832 tctatagcctcgggggttcaacttgaactcaagcgccagagcagagcttctgtggcgcgag 891
DB 721 tctatagcctcgggggttcaacttgaactcaagcgccagagcagagcttctgtggcgcgag 780
QY 892 agcgcgctcccgggcgctgtgcccctgtgcaagctcagccttcacagaaagctcagctgt 951
DB 781 agcgcgctcccgggcgctgtgcccctgtgcaagctcagccttcacagaaagctcagctgt 840

```

```

QY 952 tgcgagagctcattcatcaccaccccgagctgagatcgtgacagcccgccactgtctggaaaaa 1011
DB 841 tgcgagagctcattcatcaccaccccgagctgagatcgtgacagcccgccactgtctggaaaaa 900
QY 1012 cctcttaacaatccatgtgacattgagcagcattgtcggggattctgagacaattctcatgt 1071
DB 901 cctcttaacaatccatgtgacattgagcagcattgtcggggattctgagacaattctcatgt 960
QY 1072 tctatagcggcgatatacagaatagaaaaagtgattcttccatccaaattatgacctcaag 1131
DB 961 tctatagcggcgatatacagaatagaaaaagtgattcttccatccaaattatgacctcaag 1020
QY 1132 accaagaacatgacattgtcgctgagatgaagctgcgacgaagccctctgacttcaacgaccta 1191
DB 1021 accaagaacatgacattgtcgctgagatgaagctgcgacgaagccctctgacttcaacgaccta 1080
QY 1192 gtgaaacagctgtgtctgcgcacacacacacacacacacacacacacacacacacac 1251
DB 1081 gtgaaacagctgtgtctgcgcacacacacacacacacacacacacacacacacacac 1140
QY 1252 attccgggtgtggggcgccacgagagaaaggaaggaactctcagaagtgtctgaagctgtgc 1311
DB 1141 attccgggtgtggggcgccacgagagaaaggaaggaactctcagaagtgtctgaagctgtgc 1200
QY 1312 aaggtgctctcattgagacacagagatgcaacagacagatagctatgacaacactgtatc 1371
DB 1201 aaggtgctctcattgagacacagagatgcaacagacagatagctatgacaacactgtatc 1260
QY 1372 acacagcagatgactgtgtgcggtctctcgcaggggaaacgtcgtatctgtcgaagggtgac 1431
DB 1261 acacagcagatgactgtgtgcggtctctcgcaggggaaacgtcgtatctgtcgaagggtgac 1320
QY 1432 agtggagggcctctgtgtctctcgcgaagaacatattctgtgtgtgtgtgtgtgtgtgtgt 1491
DB 1321 agtggagggcctctgtgtctctcgcgaagaacatattctgtgtgtgtgtgtgtgtgtgtgt 1380
QY 1492 tgggtgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1551
DB 1381 tgggtgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
QY 1552 acggaactgattatcgaacaattgaaaggcagacggttaa 1590
DB 1441 acggaactgattatcgaacaattgaaaggcagacggttaa 1479

```

```

RESULT 12
US-08-807-151-2
; Sequence 2, Application us/08807151
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```



```

OY 342 NNDIALMCKLQKPIFNLDVAVPCLPNGMMLQPEOLCWMGKTEKGTSEVLNAKY 401
DB 948 RIISLEHCO5-Y-FDMKTTITRMICAGYESGTQSCMGSDGPLYCEKPGRWTFGLTS 1005
OY 402 LLEIETORCNSRYRD-NLIRPAMICAGFLQGNVDSGCGDGLVTSKNNIMW-LIGDTS 459
DB 1006 WSGVCFSKVLGPGVYSNVSYFVEMIKRQI 1034
OY 460 WSGSGC-AKAYRPGVGNVWFTDIYROM 487

RESULT 2
ID 097506 PRELIMINARY; PRT; 643 AA.
AC 097506;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLASMA.
RA TAKAHASHI T., KIMURA A., OKIMURA H., HAMABATA T.;
RT "Porcine liver plasma kallikrein.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB022425; BAA37147.1;
DR HSSP: P00766; 1CHG.
SO SEQUENCE 643 AA; 72227 MW; F0154450 CRC32;

Query Match
Best Local Similarity 42.7%; Score 751; DB 6; Length 643;
Matches 108; Conservative 48; Mismatches 89; Indels 8; Gaps 7;

DB 379 SLCLRGSDHSACATKANTRTVGGTDFLEMPQVSLQKLRQNLGGSTIGHQVNL 438
OY 237 SLR-CLACGVNLSSR-OSRIVGGSALPGAMPQVSLHVO-NV--HVGGSITTPWIV 291
DB 439 TAHCED-GLSDPIWRIYGLINISEITKTEPPSQYKEIITHQNYKLBSGDIALKL 497
OY 292 TAHCYKPLNNPMHWAFAGILROSPMFAGY-QVEKYISHNYDSKTKNNIDIALMKL 350
DB 498 ERLNLTDFQKPICLPSRDDTNVYVTCWGTGTEKEGIONITLQKNIPLVSNQECQ 557
OY 351 QKRLTFNDLVKPCLPKPMMLQPEQLCWTISGATEKEGKTSVLAAYLLIETQRCN 410
DB 558 KSTR-DHKIKQMCACGYKGGKADCKGSGGPLYCKYNGIMHLYGTSNGEGCARQEP 616
OY 411 SRVYVDLIRPAMICAGFLQGNVDSGCGDGLVTSKNNIMW-LIGDTSWGGCAKAYRP 470
DB 617 GVTYKVEYMDWI 629
OY 471 GVTGNVWFTDI 483

RESULT 3
ID 092319 PRELIMINARY; PRT; 1113 AA.
AC 092319;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN 4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 98429596.
RA TOMIYA Y., KIM D.-H., MAGOORI K., FUJINO T., YAMAMOTO T.T.;
RT "A novel low-density lipoprotein receptor-related protein with type II
membrane protein-like structure is abundant in heart.";

```

```

RL EMBL: 124:784-789(1998).
DR EMBL: AB013874; BAA34371.1; -.
DR HSSP: P00763; IDPO.
DR PROSITE: PS01209; LDLRA_1; 6.
KM Glycoprotein.
SO SEQUENCE 1113 AA; 122984 MW; ABAD31E CRC32;

Query Match
Best Local Similarity 39.1%; Score 702; DB 4; Length 283;
Matches 97; Conservative 50; Mismatches 91; Indels 6; Gaps 5;

DB 723 CODNELECANHE-CVPRDLCDGWDGSDSDSEMGCTLSKNGSSSLTVHRSKA-KE-H 779
OY 113 CSNSGIECDSSGTCINPMSNCDDVSHCPGEGEDNRCVRL--YQ-PNFIQVYSSQKSMH 169
DB 780 HVCADWRETLSQLACKMGAGEPSYTKL-IPQEQGQ-WLRLYPMENLNGSTLOELLV 837
OY 170 PCQDDNMENYGRACRDMGYKNMFYSSQGIYVDSGSTSPMKLNTS-AG-NVDYKKY-- 225
DB 838 YRHSCPSRSEISLCSKQDCGRRPAAAMKRIIGRTSRPGRPPQCSLOSERSGICCC 897
OY 226 HSDACSKAVVSLRCI--ACGVNLSSRSRIVGGSALPGAMPQVSLHVO-NVHVCQG 282
DB 898 VLIAKKWLTVACFCGREDADYMK-VFQINLDRPSGFMOTRF-VKTIILPRYSRAV 955
OY 283 SITTPETVTAACVE-KPLNPMHWTAFA-GILROSPFYGAGYVEKYISHPNYDSKT 340
DB 956 VDYDISVELSDIDINETSRYRPLSPPEEYLPDYCYITGNGHGNK-MPEK-LQEGE 1013
OY 341 KNNIDIALMKLQKPIFNLDVAVPCLPNGMMLQPEOLCWTISGATEKEGKTSVLAAY 400
DB 1014 VRIIPEQCS-Y-FDMKTTITRMICAGYESGTQSCMGSDGPLYCEKPGRWTFGLT 1071
OY 401 VLIETORCNSRYRD-NLIRPAMICAGFLQGNVDSGCGDGLVTSK-NNIMW-LIGDT 458
DB 1072 WSGVCFSKVLGPGVYSNVSYFVEMIKRQI 1101
OY 459 WSGSGC-AKAYRPGVGNVWFTDIYROM 487

RESULT 4
ID 095518 PRELIMINARY; PRT; 283 AA.
AC 095518;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE DL1170K4.2 (NOVEL TRYPSIN FAMILY PROTEIN WITH CLASS A LDL RECEPTOR
DOMAINS) (FRAGMENT).
GN DL1170K4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA BLAKEY S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL022514; CAI18442.1; -.
DR HSSP: P00763; IDPO.
KM Receptor.
FT NON TER.
SO SEQUENCE 283 AA; 31292 MW; E0FE50DB CRC32;

Query Match
Best Local Similarity 39.1%; Score 702; DB 4; Length 283;
Matches 97; Conservative 50; Mismatches 91; Indels 6; Gaps 5;

DB 38 CDGLOGPS-SRIYGAVSSEGEPMQASLOVGRHICGALLADRMVITPAHCFQEDSM 96
OY 243 GCVNLSSRSRIVGGSALPGAMPQVSLHVO-NVHVCSSITTPETVTAACV-EKPL 301
DB 97 ASYVLTWTFVFGKYNWSRMPGEVSKVSRLLRPYHEEDSHDYVALLDLDRPVSAAV 156
OY 302 NNPMHWTAFAFGILROSPMFAGY-QVEKYISHPNYDSKTKNNIDIALMKLQKPIFNLDV 360

```


Db 2285 TGGCTCTGGCGTGCAGTGCATTTGGATATCATGTGCACCTGCCATTCAGGCCAGG 2344
 Oy 1528 gtgtacgggaatgtgatgtatttcaggaactgcat 1562
 Db 2345 GTGTATGCCCGGTCGCCAAGGTTCAACAGATGAT 2379

RESULT 3
 PCT-US94-00616-1
 Sequence 1, Application PC/TUS9400616
 GENERAL INFORMATION:

APPLICANT:
 TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
 NUMBER OF SEQUENCES: 33
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/00616
 FILING DATE:

CLASSIFICATION:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2581 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 PCT-US94-00616-1

Query Match 6.7%; Score 115.8; DB 6; Length 2581;
 Best Local Similarity 50.9%; Pred. No. 3.4e-22;
 Matches 354; Conservative 0; Mismatches 332; Indels 9; Gaps 3;

Oy 874 aggaattgtggcgcgagagagcgctcccgaggcgctggcctggcagtgctcaagctgac 933
 Db 1688 AAGATTGTGCGAGAGAGTACTCCAGAGAGAGCGCTGGCTGGTGGTCTGTGAT 1747
 Oy 934 gtccagaagcgctccagctgtggcgagcgctccatccaccgcccgagtgatcgagcagcc 993
 Db 1748 TTGACGCGTCAACAGGTTGTGGAGCTTCTGTGACAGAGGATGGCTGGTGGC 1807
 Oy 994 gccacactgtggaagaaaccccttaacaatcatgtgcatgtgagcgcatlttgcgaggat 1053
 Db 1808 GCCCAGCTCGGTACGGGAGAAATGAGCGCTTAAGTGGAAAGCAGTGTAGCGCTG 1867
 Oy 1054 ctgagacaattctcaatg--tctcatgagcgagatccaaagtagaanaagtgattctc 1111
 Db 1868 CATATGCGATCAAAATCTGACTTCTCCTCAGATAGAACTAGAGTTGATGACCAATTTGTC 1927
 Oy 1112 at-----cgaattatgtactccaagacgaagaatgactgtgctgataagctgag 1167
 Db 1928 ATAAACCCACACACTCAATAAAGCGAAGAAACAAATGACATTGCTATATCAAGGTTCT 1987
 Oy 1168 aagcctctgacttcaacgacctagtgaaacagtgctgtcccaaccagagcatgagtg 1227
 Db 1988 ATGAAGTGAAGTACAGATTTATATACAGCTATTGTTTACGAGAAATAAATAAGTT 2047
 Oy 1228 ctgagacgaagcagctctgtgattccgggtggggggcgccagagagaagaaggag 1287
 Db 2048 TTTCCCCCAAGAAATTTGTTCTATTCTGTGCTGGGGGCGCACTTATATCAAGGTTCT 2107
 Oy 1288 acccagagagtgctgagagcgcgccagagtgcttctcatgtgagacaaagagcaagc 1347
 Db 2108 ACTGAGAGCTGACAGCAAGCTGACGTTCCCTTCTATCAATGAGAAAGTGA---A 2164
 Oy 1348 agatgtctctgacaacactgctcacacagcagcatgctgtgcccgcctccctcagagg 1407
 Db 2165 CACAGATGCCAGAAATATACATTACGAGAAATAATGCTGTGCGAGGCTATGACACAGGA 2224

Oy 1408 aacgtcattcttcgaggggtgacagtgagggcgctctgctgaacttcgaagaacatalc 1467
 Db 2225 GGGGTAAGTCTTGTCAAGGGGAGTTCAGGGCAGACACTGATGTCGAAGAAACACAGGA 2284
 Oy 1468 tgggtgctgtalagggagatacaagctgggggtctgtgctgtgcaaaagcttacagacga 1527
 Db 2285 TGGCTCTGGCGTGCAGTGCATTTGGATATCATGTGCACCTGCCATTCAGGCCAGG 2344
 Oy 1528 gtgtacgggaatgtgatgtatttcaggaactgcat 1562
 Db 2345 GTGTATGCCCGGTCGCCAAGGTTCAACAGATGAT 2379

RESULT 4
 US-08-508-448C-15
 Sequence 15, Application US/08508448C
 Patent No. 5804410
 GENERAL INFORMATION:
 APPLICANT: Kazuyoshi YAMAOKA et al.
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
 TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack
 STREET: 805 Fifteenth Street, N.W., #700
 CITY: Washington
 STATE: D.C.A.
 COUNTRY: U.S.A.
 ZIP: 20005

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 COMPUTER: IBM compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/508,448C
 FILING DATE: July 28, 1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Warren M. Cheek, Jr.
 REGISTRATION NUMBER: 33,367
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-8850
 TELEFAX:

TELEX:
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1517 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: trachea

US-08-508-448C-15
 Query Match 6.1%; Score 105.8; DB 2; Length 1517;
 Best Local Similarity 49.9%; Pred. No. 1.4e-19;
 Matches 356; Conservative 0; Mismatches 342; Indels 15; Gaps 3;

Oy 869 agagagagattgtgtggcgcgagagcgctcccgaggcgctggcctggagagtgagcc 928
 Db 612 AGCAGAGAACTCTTGGAGGAGCTGAGAGAGGAGAGCTGCGCCGAGCAATCTAGTC 671
 Oy 929 tgcacgtccagaaagctgacgtgtgagaggtccatcatcaccccgagtgagtcgtga 988

OY 415 YDNLITPAMCAGFLQGNVDSGDSGGLVTSKNNIM-WLIDTSGSCGCAKAYRGVY 473

DB 668 TRLPFRDWI 677

OY 474 GNVVFTDWM 483

RESULT 8

ID 091674 PRELIMINARY; PRT: 1524 AA.

AC 091674;

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

DE POLYPROTEIN (OVOCHYMASE).

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.

RN (1)

RP SEQUENCE FROM N.A.

RA YANG J.C., LINDSAY L.L., HEDRICK J.L.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1273-1524 FROM N.A.

RA YANG J.C., HEDRICK J.L.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

EMBL: U81290; AAC24717.1; -

DR EMBL: U44951; AAA91466.1; -

DR HSSP: P00763; IDPO.

DR PFAM: PF00431; CUD; 5.

DR PFAM: PF00089; trypsin; 3.

DR PRINTS: PR00722; CHYMOTRYPSIN.

KW Polypeptid.

FT CHAIN 57 308 SERINE PROTEASE.

FT CHAIN 584 817 SERINE PROTEASE.

FT CHAIN 1295 1524 OVOCHYMASE.

SO SEQUENCE 1524 AA; 167566 MW; 40D967CE CRC32;

- Query Match 18.4%; Score 677; DB 13; Length 1524;

Best Local Similarity 41.5%; Pred. No. 1,44e-147;

Matches 103; Conservative 54; Mismatches 79; Indels 12; Gaps 12;

DB 570 CGNAPTPKMLPRIVGEGEASPNMPVOIFLRTFCEGAIIISQWILTAHCT-RA 628

OY 243 CGVN-LNSR-QSRIVGESALPGAMPQVSLHVNHVCGSITTEPMIVTAHCVEKP 300

DB 629 -APPSTVTVIAGD-NHR-MLNESTEQRNKTIRHNVSEFYNDIALYLEEPDLN 665

OY 301 LNNPMHMTAAGILROSFMFYGAGYOV-E-KVIS-HPNYDSTRKNDIALMLKQPLTFN 357

DB 686 DFNVPYCLPEPEVLTLPASVCVTGNGNTAEDGQAPALGLOLPLIDSIICNTSY-YSG 744

OY 358 DLVKKPCLPFGMLPQPEOLCWSMGATEKRTSEVLAAVLLIETORCSRYIYDN 417

DB 745 ELTDHMLCAGFPSSKEDACOGSDGSGGLVLCQNEKQFSYIGLVSGEGRVSQGVYTK 804

OY 418 LITPAMICAGFLQGNV-DSGDSGGLV-TSKNNIMWLGDTSGSCGCAKAYRGVYGN 475

DB 805 VRLFTWI 812

OY 476 VNVFTDWM 483

RESULT 9

ID 088781 PRELIMINARY; PRT: 297 AA.

AC 088781;

DT 01-NOV-1998 (TEMBLrel. 08, Created)

DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

DE SERINE PROTEASE PRECURSOR (FRAGMENT).

GN BSP2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-FISHER; TISSUE-BRAIN;

RX MEDLINE; 98389725.

RA DAVIES B.J., PICKARD B.S., STEEL M., MORRIS R.G., LATHE R.;

RT "Serine proteases in rodent hippocampus."

RL J. Biol. Chem. 273:23004-23011(1998).

DR EMBL: AJ005642; CA06644.1; -

DR HSSP: P00763; IDPO.

DR PFAM: PF00089; trypsin; 1.

KW Signal: Protease; Serine protease.

FT NON_TER 1 1

FT SIGNAL <1 23 POTENTIAL.

FT CHAIN 24 297 SERINE PROTEASE.

SO SEQUENCE 297 AA; 32086 MW; 1B072619 CRC32;

- Query Match 17.0%; Score 625; DB 11; Length 297;

Best Local Similarity 36.3%; Pred. No. 3,42e-133;

Matches 87; Conservative 60; Mismatches 86; Indels 7; Gaps 5;

DB 38 NRYVGEDSDADQMPWIVSILKNGSHHCAGSLTNRNVVSAACFSSNMKPSYVLIG 97

OY 253 SRIVGESALPGAMPQVSLHVNHVCGSITTEPMIVTAHCVEKPLNNPMHTAFAG 312

DB 98 AMRLNGPSPQKAVGLASVLPFRYRKETADIALVRLERPIQSERILPICLPDSSV 157

OY 313 ILKQSEMF-YGAGYQVEKYSHPNYDSK-TKNNDIALMLKQKFLTEPNDLVKPCLPNPM 370

DB 158 HLEPNTNCWAGSGSDGVPLRPOTLQKVPRIIDPELCKSLYRAGAGEITEDMC 217

OY 371 MLEPDLCTWISGAGTEEGK-T-SEVLAAKAVLLIETORCSRY--VYNDLITPAMIC 425

DB 218 AGYLEGRKDACLDGSGPLMCQYDDMLLTGLISMGEGCAERNRPVYSLAHPVOR 277

OY 426 AGFLQGNVDSGDSGGLVTSKNNIMWLGDTSGSCGCAKAYRGVYGNVFTDWMIR 485

RESULT 10

ID 043342 PRELIMINARY; PRT: 271 AA.

AC 043342;

DT 01-JUN-1998 (TEMBLrel. 06, Created)

DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

DE SP001LA (FRAGMENT).

GN SP001LA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RN (1)

RP SEQUENCE FROM N.A.

RA RICKÉ D.O., BRUCE D., MUNDT M., DOGGETT N., MUNK C., SAUNDERS E.,

RA ROBINSON D., JONES M., BUCKINGHAM J., CHRISTEN L., THOMPSON S.,

RA GOODMAN L., BRYANT J., TESMER J., MEINKE L., LONGHIRE J., WHITE S.,

RA UENG S., TAYLOR O., CAMPBELL C., FANCERT J., DEVERN L.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA RICKÉ D.O., MAGNER R.P., MUNDT M.O.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC003965; AAB93671.1; -

DR HSSP: P00763; IDPO.

DR PFAM: PF00089; trypsin; 1.

FT NON_TER 1 1

SO SEQUENCE 271 AA; 29307 MW; FB404ADB CRC32;

- Query Match 16.6%; Score 608; DB 4; Length 271;

Best Local Similarity 37.4%; Pred. No. 1,64e-128;

Matches 89; Conservative 54; Mismatches 88; Indels 7; Gaps 5;

DB 2 NRYVGEDSDSDSWPWTIVISQKNGTHHCAGSLTNRNVVTAACFSSNMKPSYVLIG 61

OY 253 SRIVGESALPGAMPQVSLHVNHVCGSITTEPMIVTAHCVEKPLNNPMHTAFAG 312

```

GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 901 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: trachea
US-08-508-448C-9

Query Match      6.0%; Score 103.8; DB 2; Length 901;
Best Local Similarity 50.4%; Pred. No. 4,1e-19;
Matches 342; Conservative 0; Mismatches 322; Indels 15; Gaps 3;

```

```

QY 1263 gggggccaccgaggaagaagaagcctcagaagtgcgaagctgcgaaggtgctct 1322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 360 GGCGCTCAAGATATGCTGGCCACAGTCCAGAGCTAAGGCAAGACAGGTGAGAT 419
QY 1323 catggaacagagatgcaacagcagatagtctctgacaacctgaacacagcct 1382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 420 AATAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 479
QY 1383 gacgtgcccggctcctcgaagggaagctgcttcgcaagggtgagagagagc 1442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 480 GCTGTGCTGAGAGTACCTCAAGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 539
QY 1443 tctggtcacttcga--gaacatactcgtgctgataaggagatacaagctggttc 1499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 540 ACTAGTACAAAGAAAGACTCAAGCGGCTTGTATGTTAGTGAAGTGAAGTGAAGT 599
QY 1500 tggctgtgcaagcttacagacagagagtgctcgggaatgtatgttcaagagacg 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 600 TCAGTGTGCGCTGCCGGAATAAGCAGAGAGTGTATCTCGAGTGAACAGCCTTACTG 659
QY 1560 gattatcgacaaatgag 1578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 660 GATTAGCAACAACCTGG 678

RESULT
US-07-882-202A-3
Sequence 3, Application US/07882202A
Patent No. 5374617
GENERAL INFORMATION:
APPLICANT: Morrissey, James H.
APPLICANT: Comp, Philip C.
TITLE OF INVENTION: Treatment of Bleeding with Modified
TITLE OF INVENTION: Tissue Factor in Combination with FvIIa
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Richards, Medlock & Andrews
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: Texas
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/882,202A
FILING DATE: 13-MAY-1992
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hansen, Eugenia S.
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: OMRF B34290
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1440 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1433

```

```

OY 417 NLTTPMATICGFLQGNVDSGSGGGLVTSKNIMWLIGDTSNGSCACARPRGYGNV 476
DB 321 SRFVTWIEGVARN 334
OY 477 MVFTDWIYROMRAD 490

RESULT 14
ID 015146 PRELIMINARY; PRT; 810 AA.
AC 015146:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PLASMINOGEN PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RA BROWNE M.J., CHAPMAN C.G., DODD I., CAREY J.E., LAWRENCE G.M.,
RA MITCHELL D., ROBINSON J.H.;
RL Fibriolysis 0:0-0(0).
DR EMBL: M74220; AAA36451.1; -.
DR HSSP: P00747; 1PKR.
DR PFAM: PF00051; trypsin; 5.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PRO0018; KRINGLE.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
KW Signal; zymogen.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 810 PLASMINOGEN.
SQ SEQUENCE 810 AA; 90554 MW; 7E939182 CRC32;

Query Match 16.28; Score 595; DB 4; Length 810;
Best Local Similarity 39.48; Pred. No. 6,11e-125;
Matches 100; Conservative 53; Mismatches 89; Indels 12; Gaps 12;

DB 565 FDCGKPOVEPRKCPGRVYGCVAHPHSPMPOVSLRTFRGHPFGGTLISPEWVLAHCL 624
OY 241 IAGS-VNLNSR-QSRIVGESALPGAMPQVSLHV-NVHCGGSITTEPVIYTAHCV 297
DB 625 EKS-PRSSKYVILGA-HQEVNL-EPHVQ-EIEVSLLEP-TRK-DIALKLSPAVIT 678
OY 298 EKLNNPMHTATAGILRQSEFMFYAGYQVEKYISHPNYDSKTKNDIALMKLQKPLTFN 357
DB 679 DAVIPACLPSPNVVADRTCEFTVGTGSETGTG-GAGLKEADLPVLENKVCN-RIEFLN 736
OY 358 DLVAPVCLPMPGMLOPEQLCMWISGWATEKEKTSVLAARVLLLETORCNSRYV-D 416
DB 737 GRVOSTELCAGHLAGGDSGSGGLVCFEKKDYILQGVTSNGLGCARPNKPGVYVRY 796
OY 417 NLTTPMATICGFLQGNVDSGSGGGLVTSKNIMWLIGDTSNGSCACARPRGYGNV 476
DB 797 SRFVTWIEGVARN 810
OY 477 MVFTDWIYROMRAD 490

RESULT 15
ID 09XSM2 PRELIMINARY; PRT; 273 AA.
AC 09XSM2:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE TRYPTASE (EC 3.4.21.59).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-ABOMASUM;

```

```

RA PEMBERTON A.D., MCALPESSE S.M., HUNTLEY J.F., MACKELLAR A.,
RA COLLIE D.D.S., MCWILLAN L., SCODAMORE C.L., MILLER H.R.P.;
RT "DNA sequence of sheep mast cell trypsin and its immunolocalisation
RT in lung, skin and gut in comparison with sheep mast cell proteinase-
RT 1."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y18224; CAB41989.1; -.
KW Hydrolyase.
FT CHAIN 29 273 TRYPTASE.
SQ SEQUENCE 273 AA; 30288 MW; 268A4B33 CRC32;

Query Match 15.98; Score 585; DB 6; Length 273;
Best Local Similarity 39.58; Pred. No. 3.38e-122;
Matches 98; Conservative 48; Mismatches 84; Indels 18; Gaps 14;

DB 25 QRSGITGKREA-PGRMPPOVSLRVDRQYWRHOCGSLHPQVNLTAHCHIGPELOEP-- 81
OY 251 KOSRIYGESALPGA-WPMOVSLSHV-QN-VHCGGSITTEPVIYTAHCHVKNPNH 306
DB 82 -SDFRVQLREQHLRYODRLPLISRVIPHPHYVAVENGADIALQLEBPVSRHQVPTL 140
OY 307 WTAFAGILRQSEFMFYAGY-QVEKYISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCL 365
DB 141 PPASETFPPEPSCQWVGKDVNGRPLPPYPLKQYKVIENSVCDKTY-HSGLSTDYS 199
OY 366 PNPMMLOPEQLCMWISGWATEE-KGKTSE-VLNAKVLILLETORCNSRYVDNLITP-A 422
DB 200 VPIVQEDNLACGDDGRDSCGDSGLVCKVNGTWTLAGVSWGDCAPNRPRTITRIT 259
OY 423 M-TC-AGFL-QGNV--DSCGDSGGLVTSKNIMWLIGDTSNGSCACARPRGYGNV 477
DB 260 SYLDWIRH 267
OY 478 VFTDWIR 485

```

Search completed: Thu Apr 20 00:00:33 2000
Job time : 914 secs.

OY 1216 ccaagcatgctcgcagcccaagacagctcgtgattccggtgagggccaccag 1275
 DB 1020 tctgagagagcctgctcctcctcctcctcctcctcctcctcctcctcctc 1079
 OY 1276 gagaagaggagacccctcgaagtgctgaagcgtccgaagtgcttcctcattgagacag 1335
 DB 1080 gacccctggcggccagccctcgtgagctcattgctcagctcagcccgctgattgacccag 1139
 OY 1336 agatgc-----aacagcagatatgtctatgacaaactatcacaccacccatgac 1386
 DB 1140 gacctccctgcacagctcagcagagagtgagagctccccaatattcagagatgattc 1199
 OY 1387 tctgcccgtcctcgcagggagagctgattcctcgcagagtgagagtgagagctcgt 1446
 DB 1200 tctgcccgtcctcgcagggagagctgattcctcgcagagtgagagtgagagctcgt 1259
 OY 1447 gtcactcgaagaacatactcgtgctgctgataagggatataagctgggtctcgtcgt 1506
 DB 1260 gccacccactacccggggcagctgctgactgacagggcagctgctgacggccagggctgc 1319
 OY 1507 gccaaagcttaagacccagagtgtaagggagatgtgattcctcgcagctgattat 1566
 DB 1320 gccacccgtggcgcactgctgctgctgacacagggctcctccagatgactgagtgctgcaa 1379
 OY 1567 cgacaaatgagggcaga 1583
 DB 1380 aagctcattgctcctcaga 1396

RESULT 9

US-08-321-777-3

Sequence 3, Application US/08321777

Patent No. 5504067

GENERAL INFORMATION:

APPLICANT: MORRISSEY, James H.

APPLICANT: MORRISSEY, James H.

TITLE OF INVENTION: Treatment of Bleeding with Modified

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

ADDRESS: Richards, Medlock & Andrews

STREET: 1201 Elm Street, Suite 4500

CITY: Dallas

STATE: Texas

COUNTRY: US

ZIP: 75270-2197

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/321.777

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/882202

FILING DATE: 13-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: Hansen, Eugene S.

REGISTRATION NUMBER: 31,966

REFERENCE/DOCKET NUMBER: OMRF B34290C

TELECOMMUNICATION INFORMATION:

TELEPHONE: 214-939-4500

TELEFAX: 214-939-4600

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1440 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO
 AMTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: Blood
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36..1433
 OTHER INFORMATION:
 OTHER INFORMATION: factor VII cDNA

US-08-321-777-3

Query Match 5.5%; Score 95.4; DB 1; Length 1440;

Best Local Similarity 47.9%; Pred. No. 9.4e-17;

Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

OY 865 cgcagagagagatgttgaggcggagagcgcctccgggggctgagcctgcagcagtc 924
 DB 660 ccccaagccggaattgctggggggaagcgtgctccccaagggagtgctcattgagcagtc 719
 OY 925 agccctgcaagctcagagatgcaagtgctgagagctccatcatcaccccgagtgatc 984
 DB 720 cgtgttggtgtaattgacactcagttggtgggggagccctgattcaacacactgctggtg 779
 OY 985 gtgacagccgcacactgctgagaaacaccttaacatccatgcatgagcagcatt 1044
 DB 780 gtctccgcccctcctcctgcttgcacaaatcaaaactgagagactgattccgctgctg 839
 OY 1045 gctgggagatttgagaaatcttcaatgctatgagccggaatacgaagtagaaaaagt 1104
 DB 840 ggagagacagacactcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 899
 OY 1105 attctcatcaaatatgactcgaagcagcagcagcagcagcagcagcagcagcagcagc 1164
 DB 900 atcatcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 959
 OY 1165 cgaagacctcgtacttcaagcagcagcagcagcagcagcagcagcagcagcagcagc 1215
 DB 960 caccacccgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1019
 OY 1216 ccaagcatgctcgcagcccaagacagctcgtgattccggtgagggccaccagag 1275
 DB 1020 tctgagagagcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1079
 OY 1276 gagaagaggagacccctcgaagtgctgaagcgtcgcgaagtgcttcctcattgagacag 1335
 DB 1080 gacccgtggcggccagccctcgtgagctcattgctcagcagcagcagcagcagcagc 1139
 OY 1336 agatgc-----aacagcagatatgtctatgacaaactatcacaccacccatgac 1386
 DB 1140 gacctccctgcacagctcagcagagagtgagagctccccaatattcagagatgattc 1199
 OY 1387 tctgcccgtcctcgcagggagagctgattcctcgcagagtgagagtgagagctcgt 1446
 DB 1200 tctgcccgtcctcgcagggagagctgattcctcgcagagtgagagtgagagctcgt 1259
 OY 1447 gtcactcgaagaacatactcgtgctgctgataagggatataagctgggtctcgtcgt 1506
 DB 1260 gccacccactacccggggcagctgctgactgacagggcagctgacgtgagggccagggctgc 1319
 OY 1507 gccaaagcttaagacccagagtgtaagggagatgtgattcctcgcagcagcagcagc 1566
 DB 1320 gccacccgtggcgcactgctgctgctgacacagggctcctccagatgactgagtgctgcaa 1379
 OY 1567 cgacaaatgagggcaga 1583
 DB 1380 aagctcattgctcctcaga 1396

RESULT 10

PCT-US93-04493-3

Sequence 3, Application PC/TUS9304493

```

Db 279 AAGQALVDKCVTWTGWTQFYGGQAVYLOEARVPITISNENCSNPDYFGNQIKPKMFCA 338
OY 367 NGIMLQAPQLQWISGMSGATEEGKTSFVLMNAKVIILFQRCNSRYVDNLTPAMICA 426
Db 339 GTPGGIDACGDSGSHFPCEDRISTSRRLCGIYSNGTGCALAKPPVYTFKIDFREM 398
OY 427 GFLQGVNDSGDSGGPLV-TSR--NNI--WMLIGTDSGSGCAKAYRPGYGNVWFETDM 482
Db 399 IFOAIKTHS 407
OY 483 IYROMRADG 491

RESULT 11
ENTRY PLMS #type complete
TITLE plasmin (EC 3.4.21.7) precursor - mouse
CONTAINS angiotensin; plasminogen
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change 18-Jun-1999

ACCESSIONS A38514 S48202; S48203
REFERENCE A38514
#authors Degen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
#journal Genomics (1990) 8:49-61
#title Characterization of the cDNA coding for mouse plasminogen and localization of the gene to mouse chromosome 17.
#cross-references MUID:91184812
#accession A38514
#molecule_type mRNA
#residues 1-812 ##label DEG
#cross-references GB:J04766; NID:g200402; PIDN:AAA50168.1; PID:g200403
S48202
#authors Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
#journal Eur. J. Biochem. (1994) 224:863-871
#title Characterization of the murine plasma fibrinolytic system.
#cross-references MUID:95010076
#accession S48202
#molecule_type protein
#residues 20-25 ##label LIJ
#accession S48203
#molecule_type protein
#residues 22-27 ##label LI2
#COMMENT Plasminogen is synthesized by the kidney and is present in plasma and many other extracellular fluids. Plasminogen is converted into plasmin by plasminogen activators, both plasminogen and its activator being bound to fibrin. Plasmin is inactivated by alpha-2-antiplasmin (see PIR:S47217) immediately after dissociation from the clot. In the presence of the inhibitor, the activation involves only cleavage after Arg-581, resulting in two chains connected by two disulfide bonds. Without the inhibitor, the activation involves also removal of the activation peptide.
#COMMENT Strimelysin 1 (see PIR:KCMSS1) acts on plasminogen to produce angiotensin. Together with enoxacin (see PIR:A56101, PIR:B56101), angiotensin acts to inhibit angiogenesis, and so may be useful in treating solid tumors.

FUNCTION #description dissolves the fibrin of blood clots; acts as a proteolytic factor in a variety of processes including embryonic development, tissue remodeling and tumor invasion; in ovulation it weakens the walls of the graafian follicle; also activates the urokinase-type plasminogen activator fibrinolysis
#pathway #superfamily plasmin; kringle homology; plasminogen-related protein precursor homology; trypsin homology; fibrinolysis; angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine proteinase; zymogen
FEATURE 1-96 #domain plasminogen-related protein precursor homology
1-19 #label PLPH\
20-812 #domain signal sequence #status predicted #label SIG\
#product plasminogen #status predicted #label PRO\

```

```

20-96 #domain activation peptide #status predicted #label AP\
79-466 #product angiotensin #status predicted #label AST\
97-581,582-812 #product plasmin #status predicted #label MAT\
97-581 #domain chain A #status predicted #label ACH\
103-181 #domain kringle homology #label KR1\
186-282 #domain kringle homology #label KR2\
275-352 #domain kringle homology #label KR3\
377-454 #domain kringle homology #label KR4\
481-560 #domain kringle homology #label KR5\
582-812 #domain chain B #status predicted #label BCH\
582-805 #domain trypsin homology #label TRY\
49-73,53-61,
103-181,124-164,
152-176,185-262,
188-316,206-245,
234-257,275-352,
296-335,324-347,
377-454,398-437,
426-449,481-560,
502-543,531-555,
568-687,578-586,
609-625,701-768,
731-747,758-786
#disulfide bonds #status predicted\
#cleavage_site glu-asn (stromelysin 1) #status predicted\
#binding_site carbohydrate (asn) (covalent) #status predicted\
#cleavage_site thr-val (stromelysin 1) #status predicted\
#cleavage_site arg-val (plasminogen activator) #status predicted\
#active_site His, Asp, Ser #status predicted experimental\
SUMMARY 624,667,762 #length 812 #molecular-weight 90846 #checksum 237
Query Match 17.3%; Score 634; DB 1; Length 812;
Best Local Similarity 41.4%; Pred. No. 4,10e-117;
Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;
Db 553 DVCDDPLCASASSFECKGKQVPEKPCGRHVGGCVANPSPMQISLRTFTGHCFCGT 612
OY 228 DACSSKAVYSLNCIACG-VNLSSR-QSRIVGESALPAMQVSLHQ-NV-HVCGGS 283
Db 613 LAPEWVLTAAHCLKES-SRP-EF--YKVL-GAHEEYIRGLDVE-ISVAKLLEPNR 666
OY 284 ITPEWIVTAACHVEKPLNPNHMTAFAGILRQSEFWYAGYQVKVISHPYDSKTKNN 343
Db 667 DIALTLSPATITTKVIVACLPSPRYVMADRTICYITGNGTGTGFGGR-LKEAOLPY 725
OY 344 DIALMLQPLTFNDLVKFCLEPNPQMLQPOLCWISGMSGATEEGKTSFVLMNAKVL 403
Db 726 IERNVCN-EVEYLNRRKSTELACGLAGVDSGDSGSPVCEKDKYILQGYTSMGL 784
OY 404 IERQRCNSRYV-DNLTITPAMICAGFLQGNVDSGDSGSPVTSKNNIWLIGTSMGS 462
Db 785 GCARPKNKPGVYVRSFVWIEREMNN 812
OY 463 GCAKAYRPGYGNVWFETDIYROMRAD 490

RESULT 12
ENTRY A61545 #type fragments
TITLE plasmin (EC 3.4.21.7) precursor - horse (fragments)
ALTERNATE_NAMES plasminogen
CONTAINS miniplasminogen
ORGANISM #formal_name Equus caballus #common_name domestic horse
DATE 26-Oct-1994 #sequence_revision 01-Nov-1996 #text_change 18-Jul-1997
ACCESSIONS A61545 S17527
REFERENCE A61545
#authors Schaller, J.; Rickli, E.E.
#journal Enzyme (1988) 40:63-69
#title Structural aspects of the plasminogen of various species.
#cross-references MUID:89005015

```

APPLICATION NUMBER: 08/065,725
FILING DATE: 21-MAY-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/662,920
FILING DATE: 28-FEB-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 13952-8-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 28..1420
OTHER INFORMATION: /codon_start= 28
OTHER INFORMATION: /product= "Factor VII"
US-08-475-845-1

Query Match 5.5%; Score 95.4; DB 2; Length 2422;
Best Local Similarity 47.9%; Pred. No. 1,2e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

QY 865 ccacagagcagaatgttgaggcgagagcgagcgtcccgaggcgctgagcagtc 924
DB 539 CCCCAAGCGCGAATTGTCGGGGCAAGGTGTCGCCCAAGGGAGTGTCCATGGCAGGTC 658
QY 925 agccgcagctcagagagctcagcggtgcgagagctcagatcagcccgagtgatc 984
DB 659 CTGTTGTTGTAATGAGACCTGATGTTGTGGGGGACCTGTGATCAACCATCTGGGTG 718
QY 985 gtacagagccgacacgtcgtgaaacaccttaacacatcagatgagcagcat 1044
DB 719 GTCTCGCGCGCCACTGTTCCAGCAAAATCAGAACTGAGAGAACTGATCCGGCTGCG 778
QY 1045 gcgggagatttgagacaatcttcatgttcatatgagcgagatacagtagaagaaatg 1104
DB 779 GCGGAGCAGACCTGACGAGACGACGAGGGGATGACAGACCGCGGGTGGCGCAGGTC 838
QY 1105 attctcatcaaatatagtactccaagaccagaagaacatgagcatctgctgagagctg 1164
DB 839 ATCATGCCAGCAGTACCTCCCGGGCACCAACACAGACATCCGCTGCTCGCGCTG 898
QY 1165 cagaaagcctctgacttcaagcagcagtagtgaacacagtgctgtgaccaa-----c 1215
DB 899 CACCAACCCGCTGCTCTCACTGACATGTGTGTCCTCTGCTGCTGCCCAAGCAGGATTC 958
QY 1216 ccaagcatatgtctcagcagcagaacagcctctgagattcccgagtgaggggagcagag 1275
DB 959 TCTGAGAGAGACCTGCGCTTCGTGCGCTTCTCATTTGTGTAGCGCGTGGGCCAGTGTG 1018
QY 1276 gagaagaggaagacactcagaagtgtcgaagcgctcgaagtgcttctcatatgagacag 1335
DB 1019 GACCGTGGCGCCACGGCCTGTGAGCTCATGTCTTCAAGTGGCCCGCGGTGATGCCAG 1078
QY 1336 agatgc-----aacagcagatatgtctatgacaacctgatacacacagcagatc 1386
DB 1079 GACTGCTCTCAGCAGTCAAGGAGGTGGAGACTCCCAAAATATCAGGGAGTCAATGTTTC 1138
QY 1387 tctgcagagctctcagcagggaaagctgattctcagcaggtgacagtgaggagcctctg 1446
|||||

DB 1139 TGTGCGGCTACTCGATGGCAGCAGACCTCTCCAAAGGGGACAGTGGAGCCCAAT 1138
QY 1447 gtacactcgaagaacaatatctgtgtgctgataaggagatacagctgggtgtcgtgctg 1506
DB 1199 GCCACCACTACCGGGGACAGTGTGTAAGTGTGAGCGGACATCTGAGCTGGGGCCAGGCTGC 1258
QY 1507 gccaaagcttacagaccagagagtgacgggaatgtgagtgatcttaacggagcggattat 1566
DB 1259 GCACCGTGGGCGCACTTGGGGGTACACAGGCTCTCCCACTACATGAGTGGCTGCA 1318
QY 1567 cgacaatgagggcaga 1583
DB 1319 AAGCTATGCGCTCAGA 1335

RESULT 12
US-08-327-690-1
Sequence 1, Application US/08327690
Patent No. 5817788
GENERAL INFORMATION:
APPLICANT: Berkner, Kathleen L.
APPLICANT: Petersen, Lars C.
APPLICANT: Hart, Charles E.
APPLICANT: Hedner, Ulla
APPLICANT: Bregengaard, Claus
TITLE OF INVENTION: Modified Factor VII
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
STREET: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza, Steuart Street Tower
CITY: San Francisco
STATE: CA
COUNTRY: U.S.A.
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,690
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/065,725
FILING DATE: 21-MAY-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/662,920
FILING DATE: 28-FEB-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 13952-8-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 28..1420
OTHER INFORMATION: /codon_start= 28
OTHER INFORMATION: /product= "Factor VII"
US-08-327-690-1

membrane attachment using a signal-anchor sequence.
conversion from membrane-bound to soluble forms may involve further processing. See the human entry (PIR:A56318) for an alternative version of products.
mature enteropeptidase is variously reported to contain two (heavy and light) or three (mini, heavy, and light) chains derived from a single precursor form; heavy and light chains are disulfide linked

COMPLEX

FUNCTION
#description
#pathway
#classification

cleaves propeptide from trypsinogen to produce active trypsin
intestinal digestive hydrolase cascade
#superfamily enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat homology; MAM homology; scavenger receptor cysteine-rich domain homology; trypsin homology
glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; zymogen

KEYWORDS

FEATURE
22-38
52-117
118-800
199-236
358-530
547-647
653-693
694-799
801-1035

#domain transmembrane #status predicted #label TMN\
#product enteropeptidase mini chain #status predicted
#label MCH\
#product enteropeptidase heavy chain #status predicted
#label HCH\
#domain LDL receptor ligand-binding repeat homology
#label LDL1\
#domain MAM homology #label MAM\
#domain C1r/C1s repeat homology #label C1r\
#domain LDL receptor ligand-binding repeat homology
#label LDL2\
#domain scavenger receptor cysteine-rich domain homology
#status atypical #label SRCR\
#product enteropeptidase light chain #status predicted
#label LCH\
#domain trypsin homology #label TRY\
801-1030
116,147,170,194,
233,263,264,404,
456,486,519,550,
646,698,722,741,
762,864,903,965
#binding-site carbohydrate (Asn) (covalent) #status predicted\
788-912,826-842,
926-993,957-972,
983-1011
841,892,987
#disulfide_bonds #status predicted\
#active_site His, Asp, Ser #status predicted
#length 1035 #molecular_weight 114866 #checksum 4578
#score 895; DB 1; Length 1035;
Best local Similarity 36.3%; Pred. No. 3,16e-180;
Matches 138; Conservative 84; Mismatches 141; Indels 17; Gaps 16;

DB 659 CKEDNFOCKD-GECIPLVNLCDFPHCKDSDAHCVRLFNGTTDSSGLVQFRIQSI-WH 716
QY 113 GNSNGECSSGCTCINPSMNCDSVSHCPGEGEDNRCVRLT-GP-NFI-LOVYSQKRSWH 169
DB 717 VACAEMWTQISDVCOLGLGTG-NSSVPTFS-TGGGPFVNTAF-NSLSILT-PSQ 772
QY 170 PWCODHNEWGAACBDMKYNKFSSOGIVDSGSTSMKLTSGANVDIYKKHYSDA 229
DB 773 CLEDSILLQCNKSCGKLVTOEVSPIKIVGSDSREGANPVVALYFDDQVCGASLVS 832
QY 230 GSKKAVVSLRCI--ACGVNLSSRSQ-RIVGESALPGAMPQVSLHVQNVHVGCGSIIT 286
DB 833 KDLVLSAARHYGRNNE-PSKRAVILGLHMASNLTPQITRLIDQVIVPNHNRKRN 891
QY 287 PEMIVYPAARCV-EKPLNPNHMTAFAGILKROSFHFYGA-GYQ-VEKVISPNYDSKTKN 343
DB 892 DIAMHLEKVNVTDYOPICLPEENOVFPGRICISAGCALIYOGSTADVLQADAVPL 951
QY 344 DIALMKLOKRLTNDLVKPCICLPNGMLDPEQICWISGNGATBEKKTSEVLNAKVL 403
DB 952 LSNKCOQO-MPEYNTENNVCAGYENGVDSCGDSGGLPKCEKRNRLIAGYTSYQ 1010
QY 404 IETORCSRYVNLITLPAICAGFLQGNVDSGDSGLPLVSKNIMWLIGDTSWGSG 463

DB 1011 CALPNRPVYARVPRFTWMI 1030
QY 464 CAKAYRPGYGVNMVETDWI 483

RESULT 2
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE

A56318 #type complete
enteropeptidase (EC 3.4.21.9) precursor - human
enterokinase
#formal_name Homo sapiens #common_name Man
19-May-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
A56318; B43090

ACCESSIONS
REFERENCE
#authors
#journal
#title

Kitamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry (1995) 34:4562-4568
cDNA sequence and chromosomal localization of human enterokinase, the proteolytic activator of trypsinogen.
#accession A56318
#molecule_type mRNA
#residues 1-1019 #label KIT
#cross-references GB:009860; NID:9746412; PIDN:AAC50138.1; PID:9746413
A43090
#authors
#journal
#title

Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. (1994) 91:7588-7592
Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains.
#cross-references NUID:94329561
#accession B43090
#molecule_type mRNA
#residues 749-1019 #label KIT2
#cross-references GB:009860
#residues 749-1019 #label KIT2

COMMENT
The mechanism of association with the membrane of the intestinal brush border is unknown. The sequence is compatible with type II membrane attachment using a signal-anchor sequence (as annotated below) or with amino-terminal myristoylation of the heavy chain.

GENETICS
#gene
#cross-references GDB:384083; OMIM:226200
#map_position 21q21-21q21

COMPLEX
Mature enteropeptidase is variously reported to contain two (heavy and light) or three (mini, heavy, and light) chains derived from a single precursor form; heavy and light chains are linked by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involves further processing. See the pig or bovine entries (PIR:A53663, PIR:A43090) for an alternative version of products.

FUNCTION
#description
#pathway
#classification

cleaves activation peptide from trypsinogen to produce active trypsin
intestinal digestive hydrolase cascade
#superfamily enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat homology; MAM homology; scavenger receptor cysteine-rich domain homology; trypsin homology
glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen

KEYWORDS

FEATURE
1-784
22-38
184-221
342-504
526-631
643-677
678-783

#product enteropeptidase heavy chain #status predicted
#label HCH\
#domain transmembrane #status predicted #label TMN\
#domain LDL receptor ligand-binding repeat homology
#label LDL1\
#domain MAM homology #label MAM\
#domain C1r/C1s repeat homology #label C1r\
#domain LDL receptor ligand-binding repeat homology
#label LDL2\
#domain scavenger receptor cysteine-rich domain homology
#status atypical #label SRCR\
#status atypical #label SRCR

OY 773 tctataaaaaactgtc 787
 DB 15 TTTATATAAAACTGT 1

RESULT 5
 T79127
 ID T79127 standard; cDNA to mRNA; 699 BP.
 AC T79127;
 DT 08-OCT-1997 (first entry)
 DE Human serine protease 60 (SP60) cDNA.
 KM Human; colon carcinoma; COLO 201; cell line; serine protease; SP60;
 OS screening; inhibitor; treatment; disease; ds.
 PA Homo sapiens.
 FH Key Location/Qualifiers
 FT mat_peptide 1..699
 FT /tag= a
 FT /product= serine_protease
 PN J09149790-A.
 PD 10-JUN-1997.
 PE 24-JUL-1996; 212196.
 PR 29-SEP-1995; JP-275105.
 PA (SUNR) SUNTORV LTD.
 DR WPI: 97-357902/33.
 P-PSDB: W22986.
 PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
 PT - useful to screen for specific inhibitors, e.g. to search for, or
 PT study agent for treatment of various diseases
 PS Claim 2; Pages 11-12; 16pp; Japanese.
 CC The present sequence encodes the human colon carcinoma COLO 201
 CC cell line derived serine protease 60 (SP60), which can be used to
 CC screen for specific inhibitors, e.g. to search for, or study an
 CC agent for the treatment of various diseases.
 SQ Sequence 699 BP; 157 A; 190 C; 209 G; 143 T;

Query Match 7.9%; Score 137; DB 1; Length 699;
 Best Local Similarity 53.7%; Pred. No. 5,4e-29;
 Matches 381; Conservative 0; Mismatches 310; Indels 18; Gaps 4;

OY 878 tctgtggcgagagagcggtcccgaggcgctgtgcccctgagctcagcctgcagctcc 937
 DB 2 TGTGTGGGTGGGAGGAGGCGCTGTGTGATCTTGGCTTGGCAGGTCACATCCAGTACG 61
 OY 938 agaaagctcagcgtgtgctgagagctcatcatcaccaccccgagtgatgagacagcgccc 997
 DB 62 ACAAAACAGCACGCTGTGTGGAGGAGCATCTGACCCCACTGGGTCTCACGCGACCC 121
 OY 998 actgctgtgaaaaaacctcttaacaatcatalgcatgagcagcattgtcggtgatttga 1057
 DB 122 ACTGCTTCAGGAACAT-----ACCGATGTGTCAACTGGAAAGTGGCGCAGGCTCA 174
 OY 1058 gacaaacttcatgtctctatgagcgagatcaccaagtagaaaaagtattctcatccaa 1117
 DB 175 GACAAACTGGCAGCTTCCCATCCCTGCGTGTGGCCAAAGATCATCATTTGAATCAAC 234
 OY 1118 attatgacctccaagaagcaaaatagatcatgtcgctgtatgaaagctgcagaagccttga 1177
 DB 235 CCCATTGTACC-----CCAAAGACAATGATGATGCGCTCTCATGAAAGCTGATCCACTGA 289
 OY 1178 ctctcaagcactagtgaaacagatgtgtctgcccacacagcagatgagctgcagcag 1237
 DB 290 CTTTCTCAGGCGACAGTCAAGCCCATCTGTGCTCTTCTTTATGAGGAGCTCACCTCCAG 349
 OY 1238 aacagctctctgtgattcttcgggtgaggggagcagcagagagaa---aggaaagacctcag 1294
 DB 350 CCACCCCACTCTGATCATTTGGATGGGCTTACGAAGACAGATGAGGAGAGATGTCTG 409
 OY 1295 aagctctgaaagctgcagaaagtgctctcatgtgagacaaagatgcaacagaaagatag 1354
 DB 410 ACATACTGCTTGAGGCGGTGATTCAGAGTATTGACAGCACACAGGTCGAATGACAGAGATG 469
 OY 1355 tctatgacaactgatacaccaagcagatgctgtcgcgcttccgcaggggaaagctcg 1414

DB 470 CGTACCAGGGGAGTACCAGGAAGATGATGTGTGACGACATCCCGAAGGGCTGTGG 529
 OY 1415 attcttgcagggtgacagtgagggcctctgtctacctcgaagaacaatattctgttgc 1474
 DB 530 ACACCTGCCAGGGGTGACATGTGTGGGCCCCCTG---ATGTACCAATGTGACAGTGGCAGT 586
 OY 1475 tgaatagggtatacaagctgggtgtctgtgtgcacaaagcttacaagcagagagtacg 1534
 DB 587 TGTGTGGCATCGTTAGCTGTGCGGTATGCTGCGGGGCGCCGACACCCAGAGATATACA 646
 OY 1535 ggaatgtgtgtatttcaagcagctgattatcgaacaattggggagaga 1583
 DB 647 CCAAGCTTCACGCTATCTCACTGATCTCAACATGCTCGAAGGCTGA 695

RESULT 6
 X15134
 ID X15134 standard; cDNA; 1605 BP.
 AC X15134;
 DT 21-APR-1999 (first entry)
 DE cDNA encoding mouse serine protease called hepsin.
 KM Mouse serine protease; hepsin; animal model; bone disease;
 KM bone disorder; skeletal disorder; osteoporosis; Paget's disease;
 KM osteitis deformans; elevated bone alkaline phosphatase level; ss.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT CDS 35..1285
 FT /tag= a
 FT /product= hepsin
 PN W09854307-A1.
 PD 03-DEC-1998.
 PE 29-MAY-1998; E03199.
 PR 30-DEC-1997; US-000486.
 PR 30-MAY-1997; US-866058.
 PA (SCHD) SCHERING AG.
 PA (UNIW) UNIV WASHINGTON.
 PI Sadler JE, Wu Q;
 DR WPI: 99-070213/06.
 P-PSDB: W06812.

DR New nucleic acid functionally disrupts mouse hepsin gene - used to
 PT provide transgenic mice with abnormally elevated blood alkaline
 PT phosphatase, useful as models for bone disorders
 PS Claim 28; Fig 2; 29pp; English.
 CC The present sequence encodes a mouse serine protease called hepsin.
 CC The specification describes a mammalian cell in which expression of a
 CC gene encoding hepsin has been functionally interrupted or suppressed.
 CC The products and methods provide an animal model for bone disease,
 CC and are useful to determine effective treatment for bone and skeletal
 CC disorders such as osteoporosis, Paget's disease and osteitis
 CC deformans, especially those associated with elevated bone alkaline
 CC phosphatase levels.
 SQ Sequence 1605 BP; 316 A; 445 C; 487 G; 357 T;

Query Match 7.4%; Score 127.8; DB 1; Length 1605;
 Best Local Similarity 50.7%; Pred. No. 3.1e-26;
 Matches 380; Conservative 0; Mismatches 342; Indels 27; Gaps 2;

OY 873 cagagattgtgcygcygagagcggtcccgaggggcgcttgccttgcagagtcagctca 932
 DB 514 CGCATTTGTGGGGGGCAGAGACAGCTGTGGGAAGTGGCCGTGGCAGTCAAGCTGCG 573
 OY 933 cgtccagaagctccagctgtctgagagctcatcatcacccecgagtgatcgttacaagc 992
 DB 574 TTATGATGGGACCCACCTCTGTGGGGGCTCCCTGCTGTGGGCACTGGGTCTCATCTGC 633
 OY 993 cgcacactgcgtgysaaaaaccccttaacaatcatatgagcattgagcggcatttcggggat 1052
 DB 634 TGCATTTCTTTCCAGACGGAACCGGCTCTGTCTCGGTGGCAGATATTGCTGTGCTGC 693
 OY 1053 ttgagacaacttttca-----tgttctatggagccggaatacaagatga 1097

```

DB 951 LSNKCOQD-MPEYNITEMMCAEGIEGIDSCQDGPIMCLENNRLAGVTSFGYQ 1009
OY 404 ITQRCNSRYVDNITPMICAGFLQGNVDSQDGGPLVTSKNNIMWLIGDTSWGSQ 463
DB 1010 CALPNRPGYARVPRKTEMT 1029
OY 464 CAKAYRPGYGNVMTDWT 483

RESULT 4
ENTRY KOHUP #type complete
TITLE plasma kallikrein (EC 3.4.21.34) precursor - human
ALTERNATE_NAMES kininogenin; plasma prekallikrein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change
18-Jun-1999

ACCESSIONS A00921; A37939
REFERENCE A00921
#authors Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
#journal Biochemistry (1986) 25:2410-2417
#title Human plasma prekallikrein, a zymogen to a serine protease
#cross-references M01D:86243359
#accession A00921
#molecule_type mRNA
#residues 1-638 ##label CHU
#cross-references GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263
REFERENCE A37939
#authors McMullen, B.A.; Fujikawa, K.; Davie, E.W.
#journal Biochemistry (1991) 30:2050-2056
#title Location of the disulfide bonds in human plasma
prekallikrein: the presence of four novel domains in
the amino-terminal portion of the molecule.
#cross-references M01D:91152016
#accession A37939
#molecule_type protein
#residues 20-27;40-46; 'X',48; 'H',50; 'X',52-70; 'H',75-76; 'X',78-80;
103-113;131-140;141-143; 'S',144-146;147-159;187-193,
'X',195-198;199-200; 'X',202; 'H',219-228;229-230; 'X',
232-248;254-255; 'X',257-259;260-283; 'X',285;287-291,
'X',293-295;314-317; 'X',319-320;321-324; 'X',329-333,
'X',334-339; 'X',341-346; 'X',348-350;351; 'XMT',353;380-382,
'X',384-386;417-418; 'X',420-434; 'X',436-446;489-502,
'X',504-506;510-525;538-551;562; 'X',564-567;573; 'X',
575-576;578-583; 'X',585;592-604 ##label MCM
575-576;578-583; 'X',585;592-604 ##label MCM

COMMENT This protein, synthesized in the liver, circulates as a noncovalent
complex with high molecular weight (HMW) kininogen.
The zymogen is activated by factor XIIa, which cleaves the molecule
into a light chain, which contains the active site, and a heavy
chain, which associates with HMW kininogen. These chains are
linked by one or more disulfide bonds.
The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a
reciprocal reaction, factor XII after its binding to a negatively
charged surface. It also releases bradykinin from HMW kininogen
and may also play a role in the renin-angiotensin system by
converting prorenin into renin.

COMMENT
#cross-references GDB:127575; OMIM:229000
GENETICS
#gene GDB:KLK3
#map_position 4q35-4q35
CLASSIFICATION #superfamily coagulation factor XI; trypsin homology
KEYWORDS blood coagulation; duplication; fibrinolysis; glycoprotein;
hydrolase; inflammation; liver; plasma; serine proteinase
FEATURE
1-19 #domain signal sequence #status predicted #label STC\
20-338 #product plasma kallikrein #status predicted #label MAT\
20-390 #domain plasma kallikrein heavy chain #status predicted
#label HCN\
20-109 #domain apple repeat #label AP1\
110-199 #domain apple repeat #label AP2\
200-289 #domain apple repeat #label AP3\
291-380 #domain apple repeat #label AP4\
391-638 #domain plasma kallikrein light chain #status predicted

```

```

391-621 #label LCH\
21-104,47-77,51-57, #domain trypsin homology #label TRY\
111-194,137-166,
141-147,201-284,
227-256,231-237,
293-375,322-328,
383-503,419-435,
517-584,548-563,
574-602
127,308,396,453,
494
#disulfide_bonds #status experimental\
#binding_site carbohydrate (Asn) (covalent) #status
experimental\
#disulfide_bonds #status predicted\
#cleavage_site Arg-Ile (coagulation factor XIIa) #status
predicted\
#active_site His, Asp, Ser #status predicted
SUMMARY 434,483,578 #length 638 #molecular_weight 71369 #checksum 585
Query Match 20.7%; Score 762; DB 1; Length 638;
Best Local Similarity 42.9%; Pred. No. 6,826-148;
Matches 109; Conservative 51; Mismatches 84; Indels 10; Gaps 9;

DB 371 SLRLCTGDSNVCYTKTSRIVYGTSSWGEKMPQVSLQVKLQRIHCGSSLIQHWT 430
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 237 SLR-CIACGVNINSSROS-RIVGGSALPGAMPQVSLHVO-NV-HVCGGSIITPEWIV 291
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 431 TAHCDFDGLPDQVR-I-YSGILNLSIDITKDPFQKIEIITHQVYKESGNHIALIK 488
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 292 TAHCYER-PLNPNHWTAFAGILRQSFYFAGY-QVAKVISHRYDSEKTNNDALMK 349
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 489 LQAPLTYEFQKICLPSKDSSTIYNCWVTGWGFSKKEGTONILOKVINPLYTNEEC 548
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 350 LQPLTFNDLVKPVCLPNPGMMIQPEQLCMISGWGATEKSEVLNAKYLLETQRC 409
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 549 QKRYQ-DYKITQKAVAGKREGKDKCKDGGPLVCKNGMRLVGTISWEGCARRO 607
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 410 NSRYVDNITPMICAGFLQGNVDSQDGGPLVTSKNNIMWLIGDTSWGSQAKAYR 469
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 608 PGYTVKAEYMDWI 621
OY 470 PGYGNVMTDWT 483

RESULT 5
ENTRY KOHUP #type complete
TITLE plasma kallikrein (EC 3.4.21.34) precursor - rat
ALTERNATE_NAMES Fletcher factor; kininogenin; serum kallikrein
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
18-Jun-1999

ACCESSIONS A39180; A33320; S06851; I53041; S06852
REFERENCE A39180
#authors Beublen, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mdikey, M.;
Chretien, M.; Seidan, N.G.
#journal Biochemistry (1991) 30:1628-1635
#title Gene structure and chromosomal localization of plasma
kallikrein.
#cross-references M01D:91129236
#accession A39180
#molecule_type DNA
#residues 1-638 ##label BEA
#cross-references GB:J05315
#note the authors translated the codon GAG for residue 81 as
Gln
REFERENCE A33320
#authors Seidan, N.G.; Ladenheim, R.; Mdikey, M.; Hamelin, J.;
Lutalla, G.; Rougeon, F.; Lazure, C.; Chretien, M.
#journal DNA (1989) 8:563-574
#title The cDNA structure of rat plasma kallikrein.
#cross-references M01D:90091743
#accession A33320
#status not compared with conceptual translation

```

ID	Q70104 standard; cDNA; 2581 BP.	
AC	Q70104.	
DT	08-MAR-1995 (first entry)	
DE	Bovine enterokinase.	
KW	Enterokinase; EK; heavy chain; light chain; catalytic domain;	
KW	digestive disorder; cleavage; fusion protein; trypsinogen;	
KW	trypsin; enzyme; PACE gene; ss.	
OS	Bos taurus.	
FH	Key	
FT	misc-feature	Location/Qualifiers
FT		1..1690
FT		/*tag= a
FT		/note= "C-terminal portion of heavy chain
FT		(non-catalytic domain)".
FT	misc-feature	1691..2398
FT		/*tag= b
FT		/note= "light chain (catalytic domain)"
FN	MO9416083-A.	
PD	21-JUL-1994.	
PF	13-JAN-1994.	
PR	15-JAN-1993; US-005944.	
PA	(GENY) GENETICS INST INC.	
PI	Lavallie ER;	
DR	WPI; 94-249229/30.	
DR	P-PSDB; R57283.	
PT	New nucleic acid encoding enterokinase activity - and related	
PT	vectors, host cells, expression products and antibodies are	
PT	useful in treating digestive disorders and for cleaving fusion	
PT	proteins	
PS	Claim 6: Page 26-27; 50pp; English.	
CC	The enterokinase (EK) (or the EK gene when used in gene therapy) is	
CC	used to treat digestive disorders associated with low EK activity	
CC	(esp. inability to process trypsinogen to trypsin). For cleaving	
CC	fusion proteins, recombinant EK catalytic domain is much more	
CC	efficient than the native two-chain holoenzyme and is not	
CC	contaminated by other proteolytic enzymes. For expression of	
CC	recombinant EK, the 1691-2398 DNA fragment was fused to the 3'-end of	
CC	the signal peptide and pro-region of the human PACE gene. The prod.	
CC	could be expressed in CHO cells to produce a chimeric prod. from	
CC	which the pro-region as cleaved by endogenous PACE, providing mature	
CC	EK catalytic domain.	
SO	Sequence 2581 BP; 797 A; 509 C; 575 G; 700 T;	
Query Match	6.7%; Score 115.8; DB 1; Length 2581;	
Best Local Similarity	50.9%; Pred. No. 8.9e-23;	
Matches 354; Conservative 0; Mismatches 332; Indels 9; Gaps 3;		
OY	874 aggattgtggcgcgcggaagcgcgctcccggyggcccttgccgtgcaggtcagctgcac	933
DB	1668 AAGATTGTCGGAGGAGTACGTCCAGAGAGAAGAGCGCTGGCCTTGGGTGCTGTGTAT	1747
OY	934 gtccgaagcgtcccaagctgtgscgaaggtccatcatcaccccgagttgtagctgcagcc	993
DB	1748 TTCGACGATCAACAGAGCTGCGGACCTTCTCTGGTGAAGGAGGATTTGGCTGTGCGCC	1807
OY	994 gccacatcgctggaagaaacctcttaacaatccatgtgcatctggacggtcatttcgsgygaat	1053
DB	1808 GCCACCTCGTGTACGGAGAAATATGAGCGCGCTTAAGTGAAGAACAGTGTAGCGCG	1867
OY	1054 ttgagacaatccttcaag--ttctatgtaggcggaataccaagtagaanaagtgattctc	1111
DB	1868 CATATGCGATCAAAATCTGATCTCTCCTAGATAGAACTAGCTGATTATGACCAAAATTGTC	1927
OY	1112 at-----ccaattatgactccaagccaagaacaatgcatatgscgtatgtagtgcag	1167
DB	1928 ATTAACCCACACTCAATAAATGGAAGAAAGCAATGACATTGATGATGATTTGAA	1987
OY	1168 aagctctgactttcaagagacctatggaaccagtgctctcgtcccaaccaggtcagtg	1227
DB	1988 ATGGAAGTGAACACTACAGATTATATACAGCTATTGTGTTTACGAGAAATCAAGTT	2047
OY	1228 ctgcagccagaacagctctgtctgtattccgggtggggggccaccgaggaagaagggaag	1287

OY	1286	acctcaagaatgctcgaagcgctgcacaaagtgtcttcatattgagacacagagatgacaacagc	1347
DB	2108	ACTGCAGACGCTACTGTCCAAAGAACGTGACCTTCCCCTTCATCAAAATGGAAAATGTCA---A	2164
OY	1348	agatatgtctatagacaacactgtatcacaccagcaatgatctgtgcsaggtcttcctgcagggg	1407
DB	2165	CACACGATGCCAGAAATATATACATTACGGAATAATGCTGTGTGCAGCGTATGACACGAGA	2224
OY	1408	aacgtcgaatctcttcgacagtgtagacagtcgagggccctctgtctacattcgaagaacaatalc	1467
DB	2225	GCGGTATGATTTCTTGTCACGGGGGATTCAGCGGACCACTCATGTGCCAACAACAAAAACACADA	2284
OY	1468	tggctgctgatactgagagatacaaacgtcgggtctctgtcgtctgccaagaacttacagaccaga	1527
DB	2285	TGCGCTCCTGGCTGGCGGTGACGCTATTTGGATATTCATGTGTCACACTGCTTAATGCCCCAGGG	2344
OY	1528	gtgtacggagaatgtagtgtatcttcacggactgat	1562
DB	2345	GTTATGCCCCGGGTCCCAAGTTTCACAGATGGAT	2379
<hr/>			
RESULT 9			
ID	T10689	standard; DNA: 1517 BP.	
AC	T10689;		
DR	27-MAY-1996	(first entry)	
DE	Trypsin-like enzyme.		
KM	Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;		
KN	asthma; VIP; vasoactive intestinal peptide; influenza virus;		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	cds	62..1318	
FT		/tag= a	
FT	signal_peptide	62..619	
FT		/tag= b	
FT	mat_peptide	620..1315	
PN		/tag= c	
PD	AU9527248-A.		
PF	08-FEB-1996.		
PR	31-JUL-1995; 027248.		
PA	29-JUL-1994; JP-178607.		
PI	(TEIJ) TEIJI LTD.		
PI	Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;		
PI	Yamoka K, Yasuoka S;		
DR	WPI: 96-117356/13.		
PB	P-BSDb: T89435.		
PT	Nucleic acid sequence encoding trypsin-like enzyme - which digests		
PT	fibrinogen, used as expectorant in treatment of respiratory		
PT	diseases, e.g. bronchial asthma		
PS	Example 11, Page 47-49: 65pp: English.		
CC	The overlap parts of 107 bp between the sequences given in		
CC	T10698 and T10703 were identified, and thereby their identity		
CC	was confirmed. It was confirmed from the sequence analysis,		
CC	that these overlapping sequences contained a region encoding the		
CC	amino acids of the N-terminus 20 residues of the trypsin-like		
CC	enzyme isolated from the cough phlegm. The sequences were		
CC	ligated, and the desired trypsin-like enzyme gene cDNA sequence		
CC	was determined (T10689).		
SQ	Sequence 1517 BP; 478 A; 292 C; 332 G; 415 T;		
<hr/>			
Query Match 6.1%; Score 105.8; DB 1; Length 1517;			
Best Local Similarity 49.9%; Pred. No. 4.5e-20;			
Matches 356; Conservative 0; Mismatches 342; Indels 15; Gaps 3.			
OY	869	agagagagatgttgaggcgagagcgagcgctccgaggggcgctggcccctgagagtcagc	928
DB	612	ACCAAGAAATCTTGTGAGGACACTGAGGCTGAGAGAGGAGAGCTGGCCGTGGCAAGTCAGTC	671
OY	939	tgcacgtccagagclccacgctgtgcgagagctcatcatcaccgccagatgagtcgtga	988

FT	DOMAIN	106	492	EXTRACELLULAR (POTENTIAL).
FT <td>DOMAIN</td> <td>111</td> <td>150</td> <td>LDI-RECEPTOR CLASS A.</td>	DOMAIN	111	150	LDI-RECEPTOR CLASS A.
FT <td>DOMAIN</td> <td>151</td> <td>243</td> <td>SRR.</td>	DOMAIN	151	243	SRR.
FT <td>DOMAIN</td> <td>256</td> <td>492</td> <td>CATALYTIC.</td>	DOMAIN	256	492	CATALYTIC.
FT <td>ACT_SITE</td> <td>296</td> <td>296</td> <td>CHARGE RELAY SYSTEM (BY SIMILARITY).</td>	ACT_SITE	296	296	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT <td>ACT_SITE</td> <td>345</td> <td>345</td> <td>CHARGE RELAY SYSTEM (BY SIMILARITY).</td>	ACT_SITE	345	345	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT <td>ACT_SITE</td> <td>441</td> <td>441</td> <td>CHARGE RELAY SYSTEM (BY SIMILARITY).</td>	ACT_SITE	441	441	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT <td>DISULFID</td> <td>113</td> <td>126</td> <td>BY SIMILARITY.</td>	DISULFID	113	126	BY SIMILARITY.
FT <td>DISULFID</td> <td>120</td> <td>139</td> <td>BY SIMILARITY.</td>	DISULFID	120	139	BY SIMILARITY.
FT <td>DISULFID</td> <td>133</td> <td>148</td> <td>BY SIMILARITY.</td>	DISULFID	133	148	BY SIMILARITY.
FT <td>DISULFID</td> <td>244</td> <td>365</td> <td>BY SIMILARITY.</td>	DISULFID	244	365	BY SIMILARITY.
FT <td>DISULFID</td> <td>281</td> <td>297</td> <td>BY SIMILARITY.</td>	DISULFID	281	297	BY SIMILARITY.
FT <td>DISULFID</td> <td>410</td> <td>426</td> <td>BY SIMILARITY.</td>	DISULFID	410	426	BY SIMILARITY.
FT <td>DISULFID</td> <td>437</td> <td>465</td> <td>BY SIMILARITY.</td>	DISULFID	437	465	BY SIMILARITY.
FT <td>CARBOHYD</td> <td>213</td> <td>213</td> <td>POTENTIAL.</td>	CARBOHYD	213	213	POTENTIAL.
FT <td>CARBOHYD</td> <td>249</td> <td>249</td> <td>POTENTIAL.</td>	CARBOHYD	249	249	POTENTIAL.
SC <td>SEQUENCE</td> <td>492 AA;</td> <td>53847 MM; 19BC1F67 CRC32;</td> <td></td>	SEQUENCE	492 AA;	53847 MM; 19BC1F67 CRC32;	
Query Match				
Best Local Similarity		99.0%;	Score 3635;	DB 1; Length 492;
Matches		485;	Conservative	6; Mismatches 0; Indels 1; Gaps 1;
Db	1	MALNSGSPRALGPIYENHNGIYOPENPYPRAQPIVTVYVEVHPRAQYPSAPVQYARVLTA	60	
Oy	1	MALNSGSPRALGPIYENHNGIYOPENPYPRAQPIVTVYVEVHPRAQYPSAPVQYARVLTA	60	
Db	61	SNPVYCTOPKSPSGVCTSKTKKALCITLTGTFLYGALALAGLIMFMSGKCSNGSIEC	120	
Oy	61	SNPVYCTOPKSPSGVCTSKTKKALCITLTGTFLYGALALAGLIMFMSGKCSNGSIEC	120	
Db	121	DSSGTCINPSMWCQGVSHCPGEGEDENRCVRLYGFNFIOMYSQGRKSMHPYCODDNMENT	180	
Oy	121	DSSGTCINPSMWCQGVSHCPGEGEDENRCVRLYGFNFIOMYSQGRKSMHPYCODDNMENT	180	
Db	181	GRACRDMQYKNNPFSSGIGIVDDSGTSFMKLNLSAGVNDIYKRLYLSPADCCSSKAAVSLR	240	
Oy	181	GRACRDMQYKNNPFSSGIGIVDDSGTSFMKLNLSAGVNDIYKRLYLSPADCCSSKAAVSLR	240	
Db	181	GRACRDMQYKNNPFSSGIGIVDDSGTSFMKLNLSAGVNDIYKRLYLSPADCCSSKAAVSLR	240	
Oy	181	GRACRDMQYKNNPFSSGIGIVDDSGTSFMKLNLSAGVNDIYKRLYLSPADCCSSKAAVSLR	240	
Db	241	CLACGVNLNSSROSGITVGGESALPGAMPQVSLYVQNVYHVGSGSIITPEWIVTAHCVK	300	
Oy	241	CLACGVNLNSSROSGITVGGESALPGAMPQVSLYVQNVYHVGSGSIITPEWIVTAHCVK	300	
Db	301	PLNNMWHHTAARGILIRSFMYGAGIYQVAKTISHPNDSKTKNNDIALMKLQKPLTFNDL	360	
Oy	301	PLNNMWHHTAARGILIRSFMYGAGIYQVAKTISHPNDSKTKNNDIALMKLQKPLTFNDL	360	
Db	361	VKPVCLPMPGMALQPEQICWJISGMCATEBKTSSEVLNAKAVLLIETORCSRKYVDNLI	420	
Oy	361	VKPVCLPMPGMALQPEQICWJISGMCATEBKTSSEVLNAKAVLLIETORCSRKYVDNLI	420	
Db	421	TPAMICAGFLOGNDVSCOGSDGGLVTSNNIMWILIGTSMGSGCAKAYRPGVYGNVWF	480	
Oy	421	TPAMICAGFLOGNDVSCOGSDGGLVTSNNIMWILIGTSMGSGCAKAYRPGVYGNVWF	480	
Db	481	TDMITRYROMKANG	492	
Oy	481	TDMITRYROMKANG	492	
Db	480	TDMITRYROMKANG	491	
Oy	480	TDMITRYROMKANG	491	
RESULT				
ID	ENTR	2	STANDARD;	PRT; 1035 AA.
AC	ENR	BOVIN		
AC	PR6072;			
DT	01-FEB-1996	(Rel. 33, Created)		
DT	01-FEB-1996	(Rel. 33, Last sequence update)		
DT	15-DEC-1999	(Rel. 39, Last annotation update)		
DE	ENTEROPEPTIDASE	PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).		
GN	PRSS7	OR ENTK.		
OS	Bos taurus	(Bovine).		
OC	Eutelesta; Metazoa;	Chordata; Vertebrata; Mammalia;		
OC	Eutelesta; Cetartiodactyla;	Ruminantia; Pecora; Bovidae;		
OC	Bovinae;	Bos.		
RP	SEQUENCE FROM N.A.,	AND PARTIAL SEQUENCE.		

RC	TISSUE-DUODENUM;
RX	MEDLINE; 94329561.
RA	KIRAMOTO Y., YUAN X., WU O., MCCOURT D.W., SADDLER J.E.;
RT	"enterokinase, the initiator of intestinal digestion, is a mosaic
RT	protease composed of a distinctive assortment of domains.";
RL	Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN	[2]
RP	SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX	MEDLINE; 94043122.
RA	LAVALLE E.R., REHEMUTULLA A., RACIE L.A., DIBLASIO E.A.,
RT	PRENIZ C., GRANT K.L., LIGHT A., MCCOY J.M.;
RT	"Cloning and functional expression of a cDNA encoding the catalytic
RT	subunit of bovine enterokinase.";
RN	J. Biol. Chem. 268:23311-23317(1993).
RN	[3]
RP	SEQUENCE OF 801-827.
RC	TISSUE-INTESTINE;
RX	MEDLINE; 92189715.
RA	LIGHT A., JANSKA H.;
RT	"The amino-terminal sequence of the catalytic subunit of bovine
RT	enterokinase.";
RL	J. Protein Chem. 10:475-480(1991).
CC	-1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC	PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC	A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC	TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC	PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC	-1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC	TRYPSINOGEN.
CC	-1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC	MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC	-1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC	-1- PFM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC	CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC	-1- SIMILARITY: CONTAINS 2 LDB-RECEPTOR CLASS A DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC	TRYPSIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/).
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U09859; AAB40026.1; -
DR	EMBL; L19663; AAA16035.1; -
DR	PIR; A61436; A61436.
DR	HSSP; P00763; IDPO.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PS50060; MAM_2; 1.
DR	PROSITE; PS01209; LDLRA_1; 2.
DR	PROSITE; PS50068; LDLRA_2; 2.
DR	PFAM; PF00087; ldl_recept_a; 2.
DR	PFAM; PF00089; trypsin; 1.
DR	PFAM; PF00431; CUB; 2.
DR	PFAM; PF00530; SRCR; 1.
DR	PFAM; PF00629; MAM; 1.
DR	PFAM; PF01390; SEA; 1.
KW	Signal.anchor; Glycoprotein; Myristate; Hydrolase;
FT	Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
FT	CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT	CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
FT	CHAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	TRANSMEM 197 238 LDL-RECEPTOR CLASS A 1.
FT	DOMAIN 197 238


```

DR PFAM: PF00431; CUB; 2.
DR PFAM: PF00530; SRCR; 1.
DR PFAM: PF00629; MAM; 1.
DR PFAM: PF01390; SEA; 1.
KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;
Serine protease; Zymogen; Transmembrane; Repeat.
FT CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT TRANSMEM 19 47 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 182 223 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 225 334 LDL-RECEPTOR CLASS A.1.
FT DOMAIN 342 504 CUB.
FT DOMAIN 524 634 MAM.
FT DOMAIN 641 679 LDL-RECEPTOR CLASS A.2.
FT DOMAIN 671 771 SRCR.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 971 971 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 197 MYRISTATE (POTENTIAL).
FT DISULFID 184 197 BY SIMILARITY.
FT DISULFID 191 210 BY SIMILARITY.
FT DISULFID 204 221 BY SIMILARITY.
FT DISULFID 643 655 BY SIMILARITY.
FT DISULFID 650 668 BY SIMILARITY.
FT DISULFID 662 677 BY SIMILARITY.
FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
FT DISULFID 810 826 BY SIMILARITY.
FT DISULFID 910 977 BY SIMILARITY.
FT DISULFID 941 956 BY SIMILARITY.
FT DISULFID 967 995 BY SIMILARITY.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 179 179 POTENTIAL.
FT CARBOHYD 328 328 POTENTIAL.
FT CARBOHYD 335 335 POTENTIAL.
FT CARBOHYD 368 368 POTENTIAL.
FT CARBOHYD 440 440 POTENTIAL.
FT CARBOHYD 470 470 POTENTIAL.
FT CARBOHYD 503 503 POTENTIAL.
FT CARBOHYD 534 534 POTENTIAL.
FT CARBOHYD 630 630 POTENTIAL.
FT CARBOHYD 682 682 POTENTIAL.
FT CARBOHYD 706 706 POTENTIAL.
FT CARBOHYD 725 725 POTENTIAL.
FT CARBOHYD 848 848 POTENTIAL.
FT CARBOHYD 887 887 POTENTIAL.
FT CARBOHYD 909 909 POTENTIAL.
FT CARBOHYD 949 949 POTENTIAL.
SQ SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;

Query Match 24.2%; Score 890; DB 1; Length 1019;
Best Local Similarity 36.1%; Pred. No. 3,70e-210;
Matches 137; Conservative 86; Mismatches 140; Indels 17; Gaps 16;

```

```

DB 936 LSNERCQO-MPEYNTENMIGACIEGSDSCGSGPLMCOENNRFLAGVTSFGIK 994
OY 404 IETORCNSRYVNDLITPAMICAGFLOGNVDSOCGSGPLVTSKNINIMLIGDTSWMSG 463
DB 995 CALPNRPYIYARVSRTEWI 1014
OY 464 CAAKAPRGVYGNVETDWI 483

RESULT 4
ID ENTK_PTC STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE-DUODENAL MUCOSA;
RX MEDLINE; 94327548.
RA MATSUSHIMA M., ICHINOSE M., YAHAGI N., KAKEI N., TSUKADA S.,
RA MIKI K., KUROKAWA K., TASHIRO K., SHIOKAWA K., SHIMOMIYA K.,
RA UMEYAMA H., INOUE H., TAKAHASHI T., TAKAHASHI K.;
RT "Structural characterization of porcine enteropeptidase.";
RL J. Biol. Chem. 269:19976-19982(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
CC (HEAVY) CHAIN, AND A MINI CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC
DB EMBL: D30799; BAA06459.1;
DR HSSP: P00763; IDPO.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS00740; MAM; 1.
DR PROSITE: PS00740; MAM; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS01209; LDLRA_2; 2.
DR PFAM: PF00057; ldl_recept_a; 2.
DR PFAM: PF00089; trypsin; 1.
DR PFAM: PF00431; CUB; 2.
DR PFAM: PF00530; SRCR; 1.
DR PFAM: PF00629; MAM; 1.
DR PFAM: PF01390; SEA; 1.

```



```

FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 270 379 CUB.
FT DOMAIN 387 549 MAM.
FT DOMAIN 569 579 CUB.
FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 723 816 SRCR.
FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 229 242 BY SIMILARITY.
FT DISULFID 236 255 BY SIMILARITY.
FT DISULFID 249 266 BY SIMILARITY.
FT DISULFID 688 700 BY SIMILARITY.
FT DISULFID 695 713 BY SIMILARITY.
FT DISULFID 707 722 BY SIMILARITY.
FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
FT DISULFID 859 875 BY SIMILARITY.
FT DISULFID 959 1027 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1017 1045 BY SIMILARITY.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 197 197 POTENTIAL.
FT CARBOHYD 212 212 POTENTIAL.
FT CARBOHYD 373 373 POTENTIAL.
FT CARBOHYD 380 380 POTENTIAL.
FT CARBOHYD 433 433 POTENTIAL.
FT CARBOHYD 515 515 POTENTIAL.
FT CARBOHYD 579 579 POTENTIAL.
FT CARBOHYD 727 727 POTENTIAL.
FT CARBOHYD 751 751 POTENTIAL.
FT CARBOHYD 770 770 POTENTIAL.
FT CARBOHYD 791 791 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 936 936 POTENTIAL.
FT CARBOHYD 999 999 POTENTIAL.
SO SEQUENCE 1069 AA; 118735 MM; 488825A2 CMC32:

Query Match 23.08; Score 845; DB 1; Length 1069;
Best Local Similarity 36.28; Pred. No. 2,73e-197;
Matches 139; Conservative 87; Mismatches 138; Indels 20; Gaps 17;

```

```

RESULT 6
ID KAL_HUMAN STANDARD: PRT; 638 AA.
AC P03952;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KIK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN (1)
RN SEQUENCE FROM N.A.
RX MEDLINE: 86243359.
RA CHUNG D W, FUJIKAWA K, MCMULLEN B A, DAVIE E W;
RT "Human plasma prekallikrein, a zymogen to a serine protease that
RT contains four tandem repeats."
RL Biochemistry 25:2410-2417(1986).
[2]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE: 91152016.
RA MCMULLEN B A, FUJIKAWA K, DAVIE E W;
RT "Location of the disulfide bonds in human plasma prekallikrein: the
RT presence of four novel apple domains in the amino-terminal portion of
RT the molecule."
RL Biochemistry 30:2050-2056(1991).
-1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
-1- ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
-1- TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADIKININ FROM
-1- HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
-1- SYSTEM BY CONVERTING PRORENIN INTO RENIN.
-1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
-1- THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
-1- AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
-1- CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
-1- DISEASE: DEFECTS IN KIK3 ARE THE CAUSE OF FLETCHER FACTOR
-1- DEFICIENCY, A BLOOD COAGULATION DEFECT.
-1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
-1- TRYPSIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13143; AAA60153.1; -.
DR PIR: A00921; KOHRP.
DR PIR: A37939; A37939.
DR HSP: P00763; IDPO.
DR MIM: 229000; -.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS00495; APPLE; 4.
DR PFAM: PF00024; Apple; 4.
DR PFAM: PF00089; trypsin; 1.
DR Ffam: PF00089; trypsin; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Duplication; Bradykinin.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT REPEAT 20 105 APPLE 1.
FT REPEAT 110 195 APPLE 2.
FT REPEAT 200 285 APPLE 3.
FT REPEAT 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.

```


QY	896	cgctcccggggggcctgagcccttgagccaagtacagccctcagctcccaagaaagctccagctgtgagc	955
DB	841	cgctcccggggggccttgcccttgcaaggtcagccctcagctcccaagaaagctccagctgtgagc	900
QY	956	gaagctccatcaacacaccccgagctgatactgtacagcccgccacacgtcgtgaaaaaacctc	1015
DB	901	gaagctccatcaacacaccccgagctgatactgtacagcccgccacacgtcgtgaaaaaacctc	960
QY	1016	tcaacaatccatgagcatctggaacgagcatctgacggagatttgaagacaactcttcagtctct	1075
DB	961	tcaacaatccatgagcatctggaacgagcatctgacggagatttgaagacaactcttcagtctct	1020
QY	1076	atggagccggagatcccaagtacgaaaaagtattcttcattccaaatatagtaccccaagacca	1135
DB	1021	atggagccggagatcccaagtacgaaaaagtattcttcattccaaatatagtaccccaagacca	1080
QY	1136	agaaacaatgagcatctgagcttgatataagctcgcagaagccctctgacttcaagacactatgta	1195
DB	1081	agaaacaatgagcatctgagcttgatataagctcgcagaagccctctgacttcaagacactatgta	1140
QY	1196	aaccagctgtctcgcgcccaaccccaagagatgactgtcgcagccagaacacagctctgctgaattc	1255
DB	1141	aaccagctgtctcgcgcccaaccccaagagatgactgtcgcagccagaacacagctctgctgaattc	1200
QY	1256	ccgggtgtggggggccacccgagagaaagggaaagacctcagaaggtgctgtgaaacgtcgcgaag	1315
DB	1201	ccgggtgtggggggccacccgagagaaagggaaagacctcagaaggtgctgtgaaacgtcgcgaag	1260
QY	1316	tgccttcacatctggagacacagagatctgacaaagagatattctcattgaaacactgtgatacac	1375
DB	1261	tgccttcacatctggagacacagagatctgacaaagagatattctcattgaaacactgtgatacac	1320
QY	1376	cagccaatgactctgacggagctctcctcgcagggagaaacgtcagatctctgcagaggtgacagtg	1435
DB	1321	cagccaatgactctgacggagctctcctcgcagggagaaacgtcagatctctgcagaggtgacagtg	1380
QY	1436	gaagggcctcgtgcacttcgaaagaaacaatatctgtgtgctgataagggataccaagctgag	1495
DB	1381	gaagggcctcgtgcacttcgaaagaaacaatatctgtgtgctgataagggataccaagctgag	1440
QY	1496	gtctcgtgctgtgcacaaagctcttcacagaccagaggtgtctagggagatttgggttatccag	1555
DB	1441	gtctcgtgctgtgcacaaagctcttcacagaccagaggtgtctagggagatttgggttatccag	1500
QY	1556	actggaattctacacaaatgtagggcagagagctaaatccacatgctctcgtcccttgagct	1615
DB	1501	actggaattctacacaaatgtagggcagagagctaaatccacatgctctcgtcccttgagct	1560
QY	1616	cgatttacaagaaacaatgtagggctgtgttctgctcccgctgacatgattactcttagag	1675
DB	1561	cgatttacaagaaacaatgtagggctgtgttctgctcccgctgacatgattactcttagag	1620
QY	1676	atgattccagaggtcactcttattttatttaaaagagaaactgtctggc	1723
DB	1621	atgattccagaggtcactcttattttatttaaaagagaaactgtctggc	1668
RESULT 7			
PCT-US99-19655-9			
Sequence 9, Application PCT/US9919655			
GENERAL INFORMATION:			
APPLICANT: Salceda, Susana			
APPLICANT: Sun, Yongming			
APPLICANT: Reclon, Hervé			
APPLICANT: Caferkey, Robert			
APPLICANT: DIDEXUS LLC			
TITLE OF INVENTION: A NOVEL METHOD OF DIAGNOSING, MONITORING, STAGING,			
FILE OF INVENTION: IMAGING AND TREATING VARIOUS CANCERS			
FILE REFERENCE: DEX-0043			
CURRENT APPLICATION NUMBER: PCT/US99/19655			
CURRENT FILING DATE: 1999-09-01			
EARLIER APPLICATION NUMBER: 60/098, 880			

Query Match	95.2%	Score 1655.2	DB 1	Length 2479
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1660	Conservative	0	Mismatches 8	Indels 0
				Gaps 0
EARLIER FILING DATE: 1998-09-02 NUMBER OF SEQ ID NOS: 15 SOFTWARE: PatentIn Ver. 2.0 SEQ ID NO 9 LENGTH: 2479 TYPE: DNA ORGANISM: Homo sapiens PCT-US99-19655-9				
QY	56	gtcatattgaacattccagatactcatcattactctgatgctgtgttataaagaagaatg	115	
DB	1	gtcatattgaacattccagatactcatcattactctgatgctgtgttataaagaagaatg	60	
QY	116	ctttgaactcaggggtcacacacacaggtattggaaccttactatgtaaaaacatgataccac	175	
DB	61	ctttgaactcaggggtcacacacacaggtattggaaccttactatgtaaaaacatgataccac	120	
QY	176	cggaaaaccccatctccgcagacagaccactgtgtctccacatgtctcaagagtgatcccg	235	
DB	121	cggaaaaccccatctccgcagacagaccactgtgtctccacatgtctcaagagtgatcccg	180	
QY	236	ctcagtaactacccgtctcccggtgccccaagacagcccgaggggtctctgaacagagctcca	295	
DB	181	ctcagtaactacccgtctcccggtgccccaagacagcccgaggggtctctgaacagagctcca	240	
QY	296	accocgctctgtgcacagagcccaatcccatcccgaggagcagtggtcacctctaaagacta	355	
DB	241	accocgctctgtgcacagagcccaatcccatcccgaggagcagtggtcacctctaaagacta	300	
QY	356	agaaagacacgtctgcacacacttgaaacctctgagggagacctctctgtggagctgcgtccgcg	415	
DB	301	agaaagacacgtctgcacacacttgaaacctctgagggagacctctctgtggagctgcgtccgcg	360	
QY	416	ctggacctactctctgaaagttcatatgggacgaagtgctccaactctgtgataagaatgagact	475	
DB	361	ctggacctactctctgaaagttcatatgggacgaagtgctccaactctgtgataagaatgagact	420	
QY	476	ctctaaagtaacctgtatcaaaacccctctaactctgtgtgagatggcgtgtgcacaaatgccccgcg	535	
DB	421	ctctaaagtaacctgtgtatcaaaacccctctaactctgtgtgagatggcgtgtgtcaaaatgccccgcg	480	
QY	536	ggggagagagaaatcagtggtgtctgcacctctcagacagacaaactctactcctcagtgtaact	595	
DB	481	ggggagagagaaatcagtggtgtctgcacctctcagacagacaaactctactcctcagtgtaact	540	
QY	596	catctcagaagaaagttcctctgcacacctgtgtgtgccaaagacagacttgaaacgaaactacacgggc	655	
DB	541	catctcagaagaaagttcctctgcacacctgtgtgtgccaaagacagacttgaaacgaaactacacgggc	600	
QY	656	gggcggcctctcagggagacatgggctataaagaataattcttaactctcagccaaaggaataatgg	715	
DB	601	gggcggcctctcagggagacatgggctataaagaataattcttaactctcagccaaaggaataatgg	660	
QY	716	atggcagagggatcccaacacgctttatgaaacaggaaacacaaagtgcccgcaatgctgatact	775	
DB	661	atggcagagggatcccaacacgctttatgaaacaggaaacacaaagtgcccgcaatgctgatact	720	
QY	776	ataaaaaaaccttacaacacagtgatgacctctctctcctcaaaagcagtggtctctcttaacgcttla	835	
DB	721	ataaaaaaaccttacaacacagtgatgacctctctctcctcaaaagcagtggtctctcttaacgcttlt	780	
QY	836	taagcctctcgggtgtcaaatgtgaactcaagccgcgcagagacagagatgtgtggcgcgagagacg	895	
DB	781	taagcctctcgggtgtcaaatgtgaactcaagccgcgcagagacagagatcgtgtggcggtgagagacg	840	
QY	896	cgcctcccgggggcctctggacctctgcagaggtcacacccctgcagatgtccaaagacgttccacgtgtctcg	955	
DB	841	cgcctcccgggggcctctggacctctgcagaggtcacacccctgcagatgtccaaagacgttccacgtgtctcg	900	

[illegible][illegible]

Db 781 tagctcgggggtcaactctgaactcaagccgagcaagatcgtyggcgtgagagcg 840
Qy 896 cgtcccggggggctgctgcccaggtcaagctcgacgctcagaaagctcacgtgctgc 955
Db 841 cgtcccggggggctgctgcccaggtcaagctcgacgctcagaaagctcacgtgctgc 900
Qy 956 gaggctccatcaacacaccccgagtgagtcgtgacagccgcccctgctgtagaaacac 1015
Db 901 gaggctccatcaacacaccccgagtgagtcgtgacagccgcccctgctgtagaaacac 960
Qy 1016 ctacaacatcagatcgatcgagcagcattctgaggagatcttgagaaacatcttcaatctc 1075
Db 961 ctacaacatcagatcgatcgagcagcattctgaggagatcttgagaaacatcttcaatctc 1020
Qy 1076 atgagacccgagatcaagatgagaaagtgatcttctcaacaaattatgattccaaagca 1135
Db 1021 atgagacccgagatcaagatgagaaagtgatcttctcaacaaattatgattccaaagca 1080
Qy 1136 agaaacatgacatctgctgagtgagctgagagagcctctgacttcaacgacatgag 1195
Db 1081 agaaacatgacatctgctgagtgagctgagagagcctctgacttcaacgacatgag 1140
Qy 1196 aaccagtgctctgcccacacccagcgtatgctgacgagcaagacgtctgtctgatt 1255
Db 1141 aaccagtgctctgcccacacccagcgtatgctgacgagcaagacgtctgtctgatt 1200
Qy 1256 cggggtgggggggagcaagagagaaagagagagagagagagagagagagagagagag 1315
Db 1201 cggggtgggggggagcaagagagaaagagagagagagagagagagagagagagagag 1260
Qy 1316 tgcctctcatctgagacacagagatgacacagcagatagtctatgacaaactgacacac 1375
Db 1261 tgcctctcatctgagacacagagatgacacagcagatagtctatgacaaactgacacac 1320
Qy 1376 cagcactgacatctgcccgtctcctcagagggagacgtcgacttcttcagaggtgagacgtg 1435
Db 1321 cagcactgacatctgcccgtctcctcagagggagacgtcgacttcttcagaggtgagacgtg 1380
Qy 1436 gaggggctcgtgactctgaaagaaatactcgtgagcagatagagggagagagagagag 1495
Db 1381 gaggggctcgtgactctgaaagaaatactcgtgagcagatagagggagagagagagagag 1440
Qy 1496 gtctctgctgctgacaaagcttacaagacagagagtgtagaggaatgtagatcttcacg 1555
Db 1441 gtctctgctgctgacaaagcttacaagacagagagtgtagaggaatgtagatcttcacg 1500
Qy 1556 actgagattatcgacaaatgagagcagacggttaacacacatgctctgctccttgacgt 1615
Db 1501 actgagattatcgacaaatgagagcagacggttaacacacatgctctgctccttgacgt 1560
Qy 1616 cgttttaacagaaacacatgagggctggtttgtctcccgctgacatgattactctagag 1675
Db 1561 cgttttaacagaaacacatgagggctggtttgtctcccgctgacatgattactctagag 1620
Qy 1676 atgattcagaggtcactctcttctttaaacaagtgagactgtctggtc 1723
Db 1621 atgattcagaggtcactctcttctttaaacaagtgagactgtctggtc 1668

RESULT 9
US-09-342-749-29
Sequence 29, Application US/09342749
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavtjgian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TPRS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29

NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO: 29
LENGTH: 2479
TYPE: DNA
ORGANISM: Homo sapiens
US-09-342-749-29

Query Match 95.2%; Score 1655.2; DB 49; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 56 gtcatattgacattccagatcaatctatctatctgacgtctgtgataacagcaagatg 115
Db 1 gtcatattgacattccagatcaatctatctatctgacgtctgtgataacagcaagatg 60
Qy 116 ctttgaactaagggtcacacacagatctgagacattactatgaaacatgataccaaac 175
Db 61 ctttgaactaagggtcacacacagatctgagacattactatgaaacatgataccaaac 120
Qy 176 cggaaaaacccctatcccgacagccagctgtgttcccaactgtctacagaggtgacacgg 235
Db 121 cggaaaaacccctatcccgacagccagctgtgttcccaactgtctacagaggtgacacgg 180
Qy 236 ctcaagtaactacccgtcccccgtgcccagtagcccccaggggtcctgagacgcaagcttca 295
Db 181 ctcaagtaactacccgtcccccgtgcccagtagcccccaggggtcctgagacgcaagcttca 240
Qy 296 acccgcgtctgctgacgagcagcccaaatcccaatcccgagagatgtagacctcaagaacta 355
Db 241 acccgcgtctgctgacgagcagcccaaatcccaatcccgagagatgtagacctcaagaacta 300
Qy 356 agaaagcactgtgacatcaacttgacccctggggagcctctctgtggagagctgcgtgcgg 415
Db 301 agaaagcactgtgacatcaacttgacccctggggagcctctctctgtggagagctgcgtgcgg 360
Qy 416 ctggcctaactctgagaggtcatgtgagcagagtgcccaactctgagatagagtgagact 475
Db 361 ctggcctaactctgagaggtcatgtgagcagagtgcccaactctgagatagagtgagact 420
Qy 476 cctcaggtactctgacatcaacccctctaactgtgtgtagtgagctgtacacatgcccggcg 535
Db 421 cctcaggtactctgacatcaacccctctaactgtgtgtagtgagctgtacacatgcccggcg 480
Qy 536 gggagagacaggaatcggtgtgttcgctctcagagacaaactcattcattcaggttact 595
Db 481 gggagagacaggaatcggtgtgttcgctctcagagacaaactcattcattcaggttact 540
Qy 596 catctcagaggaagctctgacacccctgtgtgccaagacagctggaacgagaaactacgggc 655
Db 541 catctcagaggaagctctgacacccctgtgtgccaagacagctggaacgagaaactacgggc 600
Qy 656 gggcggcctgacaggaacatggtgctataagataatttactctagccaaggaatagtg 715
Db 601 gggcggcctgacaggaacatggtgctataagataatttactctagccaaggaatagtg 660
Qy 716 atgacagcgagatccacacagtttatgaaactgaaacaaagtgcggcaatgctgatact 775
Db 661 atgacagcgagatccacacagtttatgaaactgaaacaaagtgcggcaatgctgatact 720
Qy 776 ataaaaaacgtacacacagtgatgctgtctcctcaaaaagcagtggtcttcttaagctgta 835
Db 721 ataaaaaacgtacacacagtgatgctgtctcctcaaaaagcagtggtcttcttaagctgta 780
Qy 836 tagcctgagggtgtaactctgaactcaagccgacagagcagagatgtggtcggtgagagcg 895
Db 781 tagcctgagggtgtaactctgaactcaagccgacagagcagagatgtggtcggtgagagcg 840
Qy 896 cgtcccgagggtcgtgcccctgacagtgacagctgagacgttcagagacgtccacgtgtgag 955
Db 841 cgtcccgagggtcgtgcccctgacagtgacagctgagacgttcagagacgtccacgtgtgag 900

MPEERLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Apr 19 23:44:38 2000; Maspar time 9.00 Seconds
Tabular output not generated. 706.916 Million cell updates/sec

Title: >US-09-323-597-2
Description: (1-491) from US09323597.pep
Perfect Score: 3673
Sequence: 1 MALMSGSPALGPYENHGY.....VYGNVMTDTWYRGMADG 491

Scoring table:
PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-Issued
1:5A.COMB 2:5B.COMB 3:PCN9.COMB 4:backfiles1

Statistics: Mean 33.500; Variance 148.979; scale 0.225

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	895	24.4	798	1	US-08-200-Sequence 2, Applicatio	3.00e-71
2	895	24.4	798	3	PCT-US94-0 Sequence 2, Applicatio	3.00e-71
3	895	20.1	638	1	US-08-681-Sequence 3, Applicatio	1.19e-56
4	697	19.0	418	1	US-08-508-Sequence 25, Applicati	1.28e-52
5	691	18.8	356	2	US-08-681-Sequence 1, Applicatio	4.67e-52
6	692	18.8	855	2	US-09-027-Sequence 2, Applicatio	3.76e-52
7	678	18.5	232	1	US-08-508-Sequence 19, Applicati	7.68e-51
8	672	18.3	236	2	US-09-027-Sequence 3, Applicatio	2.79e-50
9	668	18.2	416	2	US-09-000-Sequence 2, Applicatio	6.61e-50
10	634	17.3	812	1	US-08-326-Sequence 1, Applicatio	9.85e-47
11	634	17.3	812	1	US-08-248-Sequence 1, Applicatio	9.85e-47
12	634	17.3	812	2	US-08-429-Sequence 1, Applicatio	9.85e-47
13	634	17.3	812	1	US-08-452-Sequence 1, Applicatio	9.85e-47
14	634	17.3	812	2	US-08-866-Sequence 1, Applicatio	9.85e-47
15	634	17.3	812	1	PCT-US95-0 Sequence 1, Applicatio	9.85e-47
16	634	17.3	812	1	US-08-451-Sequence 1, Applicatio	9.85e-47
17	634	17.3	812	1	US-08-605-Sequence 1, Applicatio	9.85e-47
18	634	17.3	812	2	US-08-612-Sequence 1, Applicatio	9.85e-47
19	607	16.5	270	2	US-08-978-Sequence 8, Applicatio	3.22e-44
20	598	16.3	346	4	5200340-6 Patent No. 5200340.	2.21e-43
21	597	16.3	790	1	US-08-469-Sequence 54, Applicati	2.74e-43
22	597	16.3	790	2	US-08-469-Sequence 54, Applicati	2.74e-43
23	597	16.3	791	1	US-08-643-Sequence 1, Applicatio	2.74e-43

24	597	16.3	791	2	US-08-832-Sequence 1, Applicatio	2.74e-43
25	597	16.3	791	2	US-09-131-Sequence 1, Applicatio	2.74e-43
26	597	16.3	810	4	5200340-8 Patent No. 5200340.	2.74e-43
27	597	16.3	810	1	US-07-854-Sequence 2, Applicatio	2.74e-43
28	597	16.3	810	1	US-08-147-Sequence 29, Applicatio	6.44e-43
29	593	16.1	230	2	US-08-889-Sequence 2, Applicatio	6.44e-43
30	593	16.1	230	1	US-08-379-Sequence 2, Applicatio	6.44e-43
31	593	16.1	230	1	US-08-147-Sequence 2, Applicatio	6.44e-43
32	590	16.1	814	1	US-08-750-Sequence 1, Applicatio	1.22e-42
33	587	16.0	273	2	US-08-978-Sequence 6, Applicatio	2.32e-42
34	585	15.9	230	1	US-08-456-Sequence 47, Applicati	3.57e-42
35	585	15.9	230	2	US-08-892-Sequence 47, Applicati	3.57e-42
36	585	15.9	230	1	US-08-266-Sequence 47, Applicati	3.57e-42
37	585	15.9	230	2	US-08-766-Sequence 12, Applicati	3.57e-42
38	582	15.8	273	2	US-08-978-Sequence 3, Applicatio	6.77e-42
39	577	15.7	276	2	US-09-016-Sequence 15, Applicati	1.97e-41
40	577	15.7	276	2	US-08-978-Sequence 21, Applicati	1.97e-41
41	572	15.6	267	2	US-09-016-Sequence 23, Applicati	5.73e-41
42	572	15.6	267	2	US-08-978-Sequence 18, Applicati	5.73e-41
43	572	15.6	273	2	US-08-978-Sequence 14, Applicati	5.73e-41
44	572	15.6	274	2	US-08-978-Sequence 16, Applicati	5.73e-41
45	572	15.6	274	2	US-09-016-Sequence 21, Applicati	5.73e-41

ALIGNMENTS

RESULT 1
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.

XX xxxxxx

XX Sequence 2, Application US/08200900A

CC Sequence 2, Application US/08200900A

CC Patent No. 5665566

CC GENERAL INFORMATION:

CC APPLICANT:

CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

CC NUMBER OF SEQUENCES: 38

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

CC STREET: 87 Cambridgepark Drive

CC CITY: Cambridge

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02140

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/200,900A

CC FILING DATE: 23-FEB-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Mainert, Maureen C.

CC REGISTRATION NUMBER: 31,544

CC REFERENCE/DOCKET NUMBER: GI 5201-FWC

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 876-1170 x6574

CC TELEFAX: (617) 876-5851

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 798 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 798 AA; 88937 MW; 3292434 CN;

SEQ


```

Db 157 RPYCLPARSHFEFGLCHWITGALREGEISNALOKVDVQLIPDCLSEVRY--VT 214
      :||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 361 KPYCLPBGKMLPEOLCWLISGWTGTEKGTSEVLNAKYLLETORCNSRYVDNLIT 420
      :||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 PRMLCAGYRKGKDKACGDSGSLPYCKALSGRWPLAGLVSGCGRPNTGYTRITGV 274
      :||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 421 PAMICAGFLOGNDVSCGDSGSLPY-ISKNNIMWLIIDTISGSCAKARYRGVGNWVF 479
      :||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 ISWI 278
      :|||
Oy 480 TDWI 483

RESULT 5
ID 060235; PRELIMINARY; PRT; 418 AA.
AC 060235;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE AIRWAY TRYPSIN-LIKE PROTEASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98234382.
RA YAMAOKA K., MASUDA K., OGAWA H., TAKAGI K., UEMOTO N., YASUOKA S.;
RT "Cloning and characterization of the cDNA for human airway trypsin-
RT like protease."
RL J. Biol. Chem. 273:11895-11901(1998).
DR EMBL; AB002134; BAA28691.1; -.
DR HSSP; P00750; IRTF.
DR PFAM; PF01390; SEA; 1.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
KM Protease.
SQ SEQUENCE 418 AA; 46263 MW; 0A0E0421 CRC32;

Query Match
Best Local Similarity 42.3%; Pred. No. 4.04e-153;
Matches 105; Conservative 49; Mismatches 87; Indels 7; Gaps 6;

Db 173 CGA-GPDLITLSEORILIGTAEBSGMPVOYSLKLNNAHCGGSLNNMILTAHCF-R 230
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 240 CIAGGVNLNSRQRIYVGSBGLPGAMPQVSLHYONNVHCGGSLTPEWIVTAHCYK 299
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 SNSPRQWIMNSG-STFEP--KLIMRYRNLINNNYSATHENDIALVRLNSVTFKD 287
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 300 PLNPMHMTAAGILROSEFMFGAGYOVEKYISHPNYSKTRNDIALMKLQKPLTFNDL 359
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 IHSVCLPAATONIPGSAVYTGAGDAEYAGHTVPELRQGOVRIISNDVCAHPSYNGAI 347
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 360 VKPYCLPBGKMLPEOLCWLISGWTGTEKGTSEVLNAKYLLETORCNSRYVDNLIT 419
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 LSGRLCAGVPOGVDACGDSGSLPYVDSRRLKFIYIVSGDCCGPDPRGYTRTA 407
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 420 TPAMICAGFLOGNDVSCGDSGSLPYTSKNN-IWMLIGDTISGSCAKARYRGVGNWVF 478
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 YLDWI-RQ 414
      :|||
Oy 479 FTDWIRQ 486

RESULT 6
ID 09Y495; PRELIMINARY; PRT; 571 AA.
AC 09Y495;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PLATELET FACTOR XI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98256306.
RA HSU T.C., SHORE S.K., SESHMA T., BAGASRA O., WALSH P.N.;
RT "Molecular cloning of platelet factor XI, an alternative splicing
RT product of the plasma factor XI gene."
RL J. Biol. Chem. 273:13787-13793(1998).
DR EMBL; AF045649; AAC24506.1; -.
DR PROSITE; PS00495; APPLB; 3.
SQ SEQUENCE 571 AA; 63840 MW; C25DB1A9 CRC32;

Query Match
Best Local Similarity 42.6%; Pred. No. 2.75e-152;
Matches 100; Conservative 43; Mismatches 86; Indels 6; Gaps 5;

Db 332 PRIVGTAIVRGEMPMOYTLTTSPTQRHLGSGSIIGNQWILTAHCFY-GVESPKLRV 390
      :||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 253 SRIVGESALPGAMPQVSLHYQV--HVCGGSIITPEWIVTAHCFEKLNPWMTA 309
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 YSGILNOSIEKEDTSFSGYQDEIIHDQKMAESGYDIALKLETTVNTDSORPICPSK 450
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 310 FAGILROSEFMFGAGY-OVEKVISHPNYDSKTKNNIDIALMKLQKPLTFNDLVKFCPLNP 368
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GDRNVITDCWVTGMYRKLDRKIDNTLQKAKIPLVYNECOKRYR-GHKITHKICAGY 509
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 369 GMLQPEOLCWLISGWTGTEKGTSEVLNAKYLLETORCNSRYVDNLITPAMICAGF 428
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 510 REGGRKADCGDSGSLPSCKNEVHLVGTSGGCAQREPRGYTVNYEYVDWI 564
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 429 LQGNVDSGQDSGSLPYTSKNNIMWLIIDTISGSCAKARYRGVGNWVF 483
      :| : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ID 09Y5Y6; PRELIMINARY; PRT; 683 AA.
AC 09Y5Y6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE MATRIPASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99303581.
RA LIN C.Y., ANDERS J., JOHNSON M., SANG Q.A., DICKSON R.B.;
RT "Molecular cloning of cDNA for matrilipase, a matrix-degrading serine
RT protease with trypsin-like activity."
RL J. Biol. Chem. 274:18231-18236(1999).
DR EMBL; AF118224; AAD42765.1; -.
DR PROSITE; PS01209; LDLRA_1; 2.
KM Glycoprotein.
SQ SEQUENCE 683 AA; 75626 MW; 3F788B27 CRC32;

Query Match
Best Local Similarity 40.8%; Pred. No. 9.88e-152;
Matches 102; Conservative 58; Mismatches 79; Indels 11; Gaps 9;

Db 430 CDCGLRSTFRQARVYGTDADEGEPMQVSLHALGQGHIGASLSPMVLVSAHCTIDD 489
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 243 CGVNLNS-SRQRIYVGSBGLPGAMPQVSLH-VQNVYCGGSIITPEWIVTAHCF-VE- 298
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 RGFYSDEPTQWTAFLGLHDQSORAPGQVQERRLKRIISHPEFNDTFDYDIALLEKPA 549
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 299 KPL--NPMHMTAAGILROSEFMFGAGY-OVEKVISHPNYDSKTKNNIDIALMKLQKPL 354
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 EYSSVVRPICPDASHVYRPAKAIWVTGWTGHTQYGTGATALLQGEIVINOTCEN--L 607
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Oy 355 TFNDLVKPVCLPBNPMQPEOLCWLISGWTGTEKGTSEVLNAKYLLETORCNSRYV 414
      :| : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 LPOQITPRMGCVGFSGVDSCGDSGSLPSVEADGRIFQAGVYVMDGCAQNRNKPQVY 667
      :| : : : : : : : : : : : : : : : : : : : : : : : :

```


Db 672 TCGGCTCAATTAATGCCACCACTGTGGAGGCAAGCTGATCAATTAACATGTCATCTGA 721
Qy 989 CAGCCGCCCACTGCGTGGAAACCTCTTAAACATCATGTCATGTCAGTGGCACTTGGC 1048
Db 732 CAGCAGCTCACTCTCTTCAAGAGCACTCT--AATCTCTGTGATGATGATGATGATG 788
Qy 1049 GATCTTGGAGCAATCTTCAATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1108
Db 789 GATTTCCACAACTTTCCAAACTAA-----GAATGAGAGTAAATTAATTTTAA 839
Qy 1109 CTAATCAATTAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1168
Db 840 TTTATTAACATTAATTAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 899
Qy 1169 AGCTCTGATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1228
Db 900 ACAGTGTCACTTTACCAAGATATCCATAGTGTGTGTCTCCACGCTGCTACCAAGATA 959
Qy 1229 TCGAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1288
Db 960 TTTCCACTGCTCTACTGCTTATGTAACAGATGGGCGCTCAAGAAATATGTCGCGCACA 1019
Qy 1289 CCGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1348
Db 1020 CAGTTCAGAGCTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1079
Qy 1349 GATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1408
Db 1080 CACATGATTAATGAGACCATTTCTGTCTGAAATGCTGTGTGTGTGTGTGTGTGTGT 1139
Qy 1409 ACCTGATCTCTTCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1465
Db 1140 GAGTGAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1199
Qy 1466 TCGTGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1325
Db 1200 TTTGTTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1259
Qy 1526 GAGTGAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1578
Db 1260 GAGTGAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1312

RESULT 5
US-08-508-448C-24
Sequence 24, Application US/08508448C
Patent No. 5804410
GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-508-448C-24

Query Match 6 0%; Score 104.2; DB 2; Length 696;
Best Local Similarity 49.9%; Pred. No. 2.8e-19;
Matches 352; Conservative 0; Mismatches 338; Indels 15; Gaps 3;

Qy 877 attgttgccgagagagcgctcccgaggccctggccctggcaggtcagctgcagctc 936
Db 1 ATCTTGGAGGCACTAGGCTGAGAGGAGGAGAGCTGCGCTGCACTGCTGCTGCTGCT 60
Qy 937 cagacgctcagctgctgagagctccatcatcaccacccagagtgatcgagcagcgc 996
Db 61 AATATGCGCCACACAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Qy 997 cactgctgagagagagagagagagagagagagagagagagagagagagagagag 1056
Db 121 CACTGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 177
Qy 1057 agacatcttctcagctgctgagagagagagagagagagagagagagagagagag 1116
Db 178 ACAACATTTCTTAATTA-----GAATGAGAGTAAAGAAATTTTAATTCATAC 228
Qy 1117 aattatgactcagagagagagagagagagagagagagagagagagagagagagag 1176
Db 229 AATTATTAATTTCTCACTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 288
Qy 1177 acttcaagagagagagagagagagagagagagagagagagagagagagagagag 1236
Db 289 ACCTTACCAAGAGATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 348
Qy 1237 gaacagctcgtcgtgagagagagagagagagagagagagagagagagagagagag 1296
Db 349 GCGCTACAGCTTATGTAACAGATGGGCGCTCAAGATATGCTGCGCACAGTTCCA 408
Qy 1297 gtcgtgagagagagagagagagagagagagagagagagagagagagagagag 1356
Db 409 GAGCTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
Qy 1357 tatgacaactgtatcacacacagagagagagagagagagagagagagagagagag 1416
Db 469 TATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 528
Qy 1417 tcttgcaagagagagagagagagagagagagagagagagagagagagagagag 1473
Db 529 GCATGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588
Qy 1474 ctgataagagagagagagagagagagagagagagagagagagagagagagagag 1533
Db 589 ATGTGTGGGAGTAAAGTGTGGGAGATCAAGTGTGGGAGTGTGGGAGTGTGGGAGT 648
Qy 1534 gggagagagagagagagagagagagagagagagagagagagagagagagagag 1578
Db 649 ACTGAGTGAACAGCTACCTGACTGATTAAGGAGGAGGAGGAGGAGGAGGAGGAG 693

RESULT 6
US-08-508-448C-9
Sequence 9, Application US/08508448C
Patent No. 5804410

Db 672 TGGGGCCATTAATGCCACCAGCTGGAGCGACCCGATCAATTAATGATGCTCTGA 731
 Qy 989 CACCGCCGCGCTCGTGGAAAGACCTCTAACATCATGGCATTGGAGCGCTTGC 1048
 Db 732 CACGAGCTCTCTCTTCAAGAACCACTCT--AACTCTCTGACTGATGCCACGCTG 788
 Qy 1049 GAGCTTGGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1108
 Db 789 GATTTTCCACAACTTTCCAACTAA-----GAATGAGAGTAAAGAAATTTTAA 839
 Qy 1109 CTAATCAATCTATGATGCAACTGCAATGCAATGCAATGCAATGCAATGCAATGCA 1168
 Db 840 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 899
 Qy 1169 AGCTCTGATCTTCAAGACCTGATGAAACAGTGTCTCTCTCTCTCTCTCTCTCT 1228
 Db 900 ACAAGTGCACCTTTACCAAGATATCCATAGTGTGTCTCTCTCTCTCTCTCTCT 959
 Qy 1229 TGAAGCCAGAACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1288
 Db 960 TTCCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1019
 Qy 1289 CCAAGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1348
 Db 1020 CAGTTCAGAGCTTAAGCAAGACAGTCAAGATTAATTAATTAATTAATTAATTA 1079
 Qy 1349 GATATGCTATGCAACCTGATCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1408
 Db 1080 CACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1139
 Qy 1409 AGCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1465
 Db 1140 GATTCAGAGCTTAAGCAAGACAGTCAAGATTAATTAATTAATTAATTAATTA 1199
 Qy 1466 TCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1555
 Db 1200 TTTGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1259
 Qy 1526 GAGTTCAGAGCTTAAGCAAGACAGTCAAGATTAATTAATTAATTAATTAAT 1578
 Db 1260 GAGTTCAGAGCTTAAGCAAGACAGTCAAGATTAATTAATTAATTAATTAAT 1312
 RESULT 10
 T10698
 ID T10698 standard: cDNA to mRNA: 901 BP.
 AC 26-MAY-1996 (first entry)
 DE Trypsin-like enzyme N-terminal DNA fragment from p19-33.
 KM Trypsin: fibrinogen; thrombin; expectorant; respiratory disease;
 KW asthma; VIP: vasoactive intestinal peptide; Influenza virus;
 OS protease; primer: PCR; amplification; ds.
 OS Homo sapiens.
 PN AU9527248-A.
 PD 08-FEB-1996.
 PF 31-JUL-1995: 027248.
 PR 29-JUL-1994: JP-178607.
 PA (TEIT) TEITIN LTD.
 PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;
 PI Yamada K, Yaeuoka S;
 PT Nucleic acid sequence encoding trypsin-like enzyme - which digests
 PT fibrinogen, used as expectorant in treatment of respiratory
 PT diseases, e.g. bronchial asthma
 PS Example 9, Page 44-45; 65pp; English.
 CC Example 9 describes the cloning of cDNA region encoding
 CC trypsin-like enzyme isolated from cough phlegm.
 CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8
 CC (T10695), TRX-10 (T10696) and TRX-11 (T10697).
 CC TRX-1 corresponds to from A1 to A23 of the sequence given in
 CC T10693, which is part of a cDNA encoding the trypsin-like
 CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.

CC TRX-10 is capable of annealing to the 3'-terminus of poly(A)+ RNA.
 CC TRX-11 is identical to the 5'-terminus side 19 residues of TRX-10.
 CC After amplification, plasmid p19-33 was obtained. p19-33 encodes
 CC part of the N-terminus amino acid sequence 20 residues of the
 CC trypsin-like enzyme isolated from the cough phlegm (see T10698).
 SO Sequence 901 BP; 270 A; 185 C; 205 G; 241 T;
 Query Match 6.0%; Score 103.8; DB 1; Length 901;
 Best Local Similarity 50.4%; Pred. No. 1.3e-19;
 Matches 342; Conservative 0; Mismatches 322; Indels 15; Gaps 3;
 Qy 903 GGGGCGCTGGCCCTGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 962
 Db 12 GGGAGGCTGGCCCGGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 71
 Qy 963 GATCATACCCCGAGTGGTCTGTGAGGCGCCGCTGCTGTGAGGCGCCGCTGCTGTG 1022
 Db 72 CCGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 128
 Qy 1023 TCAATGCTGAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1082
 Db 129 TCTCTGCTGAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 182
 Qy 1083 CGGATACCAAGTGAAGAAAGTGAATTCATCAATCAATCAATCAATCAATCAATCA 1142
 Db 182 --GAATGAGATGAAGAAATTTTAATTAATTAATTAATTAATTAATTAATTAAT 239
 Qy 1143 TGAATGCTGAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1202
 Db 240 TGAATGCTGAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 299
 Qy 1203 GTGTCTGCGCAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1262
 Db 300 GTGTCTGCGCAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 359
 Qy 1263 GGGGCGCTGGCCCGGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1322
 Db 360 GGGGCGCTGGCCCGGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 419
 Qy 1323 CATGAGACCCAGAGAGTGAAGACAGCAGTGTCTGTGAGGCGCCGCTGCTGTGAG 1382
 Db 420 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 479
 Qy 1383 GATCTGTGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1442
 Db 480 GCTGTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 539
 Qy 1443 TCTGTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1499
 Db 540 ACTAGTCAAGAGAGTCAAGGCGGCTTGTGTTATGTTGTTGTTGTTGTTGTTGTT 599
 Qy 1500 TGGCTGTGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1559
 Db 600 TCAAGTGTGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 659
 Qy 1560 GATTATCGAACAATGAGG 1578
 Db 660 GATTATCGAACAATGAGG 678
 RESULT 11
 T10688
 ID T10688 standard: DNA: 696 BP.
 AC 26-MAY-1996 (first entry)
 DE Trypsin-like enzyme.
 KM Trypsin: fibrinogen; thrombin; expectorant; respiratory disease;
 KW asthma; VIP: vasoactive intestinal peptide; Influenza virus;
 OS protease; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT signal_peptide 1..60

[illegible]

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Johnson, James D.
CC REGISTRATION NUMBER: 31,771
CC REFERENCE/DOCKET NUMBER: 05213-0122
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-818-3700
CC TELEFAX: 404-818-3799
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 812 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ORIGINAL SOURCE:
CC ORGANISM: Murine
CC
SQ SEQUENCE 812 AA; 90846 MW; 3411937 CN;

Query Match 17.3%; Score 634; DB 3; Length 812;
Best Local Similarity 41.4%; Pred. No. 9,85e-47;
Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;

Db 553 DYCDIPLCASASSFECGPKQVPEKPGVGGCVANPHSMPQISLRTFTGQHFQGT 612
QY 228 DACSSKAVVSLRCIACG-VNLNSR-QSRIVGGSALPGAMPQVSLHVQ-NV-HVCGGS 283
Db 613 LIAPENVLTAAHCLKS-SRP-EF--YKYL-GAHEEYINGLDVQE-ISAKLILEPNNR 666
QY 284 IITPEWIVTAAHCVEKPLNPMHWTAFAGILROSFMFYAGAYVEKYISHPNYDSKTKNN 343
Db 667 DIALTKSRPATITDKYIPACLPSPNMYVADRTICYITGNETGTGAGR-LKEAQLPV 725
QY 344 DIALMKLQKPLTFNDLVKPVCLPMPGMLOPEQLCWISGWAIEKGTISEVLNAKAVL 403
Db 726 IENKVCN-RVEYLNNRYKSTELCAGOLAGVDSGQDGGPVCFEKDKYILQGVTSMGL 784
QY 404 IETORCNSRYVY-DNLITPAMICAGFLQGHVDSGQDGGPVLVTSKNINIMWLIGDTSMS 462
Db 785 GCARPKNKPGYVYVRSFVDWIEREMNN 812
QY 463 GCAKAYRPGYVGNVFTDWIYROMRAD 490

Search completed: Wed Apr 19 23:45:02 2000
Job time : 24 secs.

 WIRENET (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Apr 19 23:22:52 2000; Maspar time 15.97 Seconds
 Tabular output not generated. 728.302 Million cell updates/sec

Title: >US-09-323-597-2
 Description: (1-491) from US09323597.pap
 Perfect Score: 3673
 Sequence: 1 MALMSGSPPAIGPYENHGY.....VYGNVMTFTWIRYQMRADG 491

Scoring table: PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: a-geneseq36
 1:geneseqp

Statistics: Mean 35.553; Variance 144.369; scale 0.246

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	695	24.4	798	1	R57283	Bovine enterokinase.	2,31e-77
2	710	19.3	233	1	W22986	Human serine protease	2,36e-58
3	697	19.0	418	1	R89435	Trypsin-like enzyme.	5,04e-57
4	691	18.8	356	1	W46917	Amino acid sequence of	2,07e-56
5	678	18.5	232	1	R89430	Trypsin-like enzyme.	4,38e-55
6	658	18.2	416	1	W96812	A mouse serine protease	4,59e-54
7	653	17.8	285	1	W77301	Amino acid sequence of	1,55e-52
8	649	17.7	241	1	W22987	Human serine protease	3,95e-52
9	634	17.3	812	1	W07585	Murine plasminogen seq	1,33e-50
10	634	17.3	812	1	R89436	Murine plasminogen.	1,33e-50
11	634	17.3	812	1	R89436	Complete mouse plasmin	1,33e-50
12	615	16.7	289	1	W77302	Amino acid sequence of	1,13e-48
13	608	16.6	271	1	W77302	Amino acid sequence of	5,80e-48
14	608	16.6	317	1	Y13391	Amino acid sequence of	7,33e-48
15	607	16.5	270	1	W64239	Gerbil homologue of mo	7,33e-48
16	606	16.5	811	1	R12933	Plasminogen muten X1	9,26e-48
17	606	16.5	812	1	R12934	Plasminogen muten X2	9,26e-48
18	606	16.5	813	1	R12935	Plasminogen muten X3	9,26e-48
19	606	16.5	813	1	R12937	Plasminogen muten X6	9,26e-48
20	605	16.5	816	1	R12946	Plasminogen muten T19	1,17e-47
21	604	16.4	297	1	W77304	Amino acid sequence of	1,48e-47
22	603	16.4	810	1	R12938	Plasminogen muten T1	1,87e-47
23	603	16.4	811	1	R12943	Plasminogen muten T13	1,87e-47

24	603	16.4	811	1	R12945	Plasminogen muten T17	1,87e-47
25	603	16.4	811	1	R12944	Plasminogen muten T14	1,87e-47
26	603	16.4	811	1	R12939	Plasminogen muten T2	1,87e-47
27	603	16.4	812	1	R12948	Plasminogen muten T5	1,87e-47
28	604	16.4	813	1	R12936	Plasminogen muten X2	1,48e-47
29	597	16.3	437	1	W51457	Human plasminogen frag	7,57e-47
30	598	16.3	546	1	R34427	Sequence of tissue pla	5,99e-47
31	597	16.3	790	1	R60519	Human 'Glu' plasminoge	7,57e-47
32	597	16.3	791	1	W34285	Human plasminogen.	7,57e-47
33	597	16.3	810	1	W31169	Plasminogen protein fo	7,57e-47
34	597	16.3	810	1	R08065	Human plasminogen from	7,57e-47
35	597	16.3	810	1	R34428	Sequence encoded by a	7,57e-47
36	596	16.2	230	1	R56471	Mutant plasmin proteas	9,56e-47
37	594	16.2	314	1	W77296	Amino acid sequence of	1,52e-46
38	594	16.2	314	1	W97116	A human eosinophil ser	1,52e-46
39	594	16.2	314	1	Y13388	Amino acid sequence of	1,52e-46
40	595	16.2	816	1	R12942	Plasminogen muten T8	1,21e-46
41	593	16.1	230	1	R56470	Mutant plasmin proteas	1,92e-46
42	593	16.1	230	1	R49031	Plasmin protease domai	1,92e-46
43	593	16.1	243	1	R54766	Serine protease domain	1,92e-46
44	593	16.1	812	1	R12940	Plasminogen muten T6	1,92e-46
45	593	16.1	812	1	R12941	Plasminogen muten T7	1,92e-46

ALIGNMENTS

RESULT 1
 ID R57283; standard; Protein: 798 AA.

AC R57283; 08-MAR-1995 (first entry)
 DE Bovine enterokinase.
 DE Enterokinase; EK; heavy chain; light chain; catalytic domain;
 KM digestive disorder; cleavage; fusion protein; trypsinogen;
 KM trypsin; enzyme; PACE gene.
 OS Bos taurus.
 FH Key
 FT domain
 FT 1..563 Location/Qualifiers
 FT /label= "heavy-chain_C-terminal
 FT /note= "non-catalytic domain"
 FT 564..798
 FT /label= "light-chain
 FT /note= "catalytic domain"
 FT domain
 FT 1..563 Location/Qualifiers
 FT /label= "heavy-chain_C-terminal
 FT /note= "non-catalytic domain"
 FT 564..798
 FT /label= "light-chain
 FT /note= "catalytic domain"
 DR M09416083-A.
 PD 21-JUL-1994.
 PE 13-JAN-1994; U00616.
 PR 15-JAN-1993; US-005944.
 PA (GEMT) GENETICS INST INC.
 PI Levallic ER.
 DR MPI: 94-249229/30.
 DR N-PSDB; 070104.
 PT New nucleic acid encoding enterokinase activity - and related
 PT vectors, host cells, expression products and antilodies are
 PT useful in treating digestive disorders and for cleaving fusion
 PT proteins
 PS Disclosure: Page 28-30; 50pp; English.
 CC The enterokinase (EK) (or the EK gene when used in gene therapy) is
 CC used to treat digestive disorders associated with low EK activity
 CC (esp. inability to process trypsinogen for cleaving fusion
 CC proteins, recombinant EK catalytic domain is much more
 CC efficient than the native two-chain holoenzyme and is not
 CC contaminated by other proteolytic enzymes. For expression of
 CC recombinant EK, the 1691-2398 DNA fragment was fused to the 3'-end of
 CC the signal peptide and pro-region of the human PACE gene. The prod.
 CC could be expressed in CHO cells to produce a chimeric prod. from
 CC which the pro-region as cleaved by endogenous PACE, providing mature
 CC EK catalytic domain.
 SQ Sequence 798 AA:

Query Match 24.4%; Score 895; DB 1; Length 798;
 Best Local Similarity 36.3%; Pred. No. 2,31e-77;
 Matches 138; Conservative 84; Mismatches 141; Indels 17; Gaps 16;
 DB 422 CKEDNFOCKD-GEICIPVNLCDGFPHCKDSDAHCVLENGTDSGLVQFRIQS1-WR 479


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query	Match	Best Local Similarity	Score 1721	DB 28	Length 3289
<p>ATTORNEY/AGENT INFORMATION: NAME: Becker, Cheryl L. REGISTRATION NUMBER: 35,441 REFERENCE/DOCKET NUMBER: 6182, US, 01 TELECOMMUNICATION INFORMATION: TELEPHONE: 847/935-1729 TELEFAX: 847/938-2623 TELEX: INFORMATION FOR SEQ ID NO: 16: SEQUENCE CHARACTERISTICS: LENGTH: 3289 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear MOLECULE TYPE: cDNA US-08-943-570-16</p>					
Query	1	99.9%	99.9%	1721	3289
Db	36	99.9%	99.9%	1721	3289
Query	61	99.9%	99.9%	1721	3289
Db	96	99.9%	99.9%	1721	3289
Query	121	99.9%	99.9%	1721	3289
Db	156	99.9%	99.9%	1721	3289
Query	181	99.9%	99.9%	1721	3289
Db	216	99.9%	99.9%	1721	3289
Query	241	99.9%	99.9%	1721	3289
Db	276	99.9%	99.9%	1721	3289
Query	301	99.9%	99.9%	1721	3289
Db	336	99.9%	99.9%	1721	3289
Query	361	99.9%	99.9%	1721	3289
Db	396	99.9%	99.9%	1721	3289
Query	421	99.9%	99.9%	1721	3289
Db	456	99.9%	99.9%	1721	3289
Query	481	99.9%	99.9%	1721	3289
Db	516	99.9%	99.9%	1721	3289
Query	541	99.9%	99.9%	1721	3289
Db	576	99.9%	99.9%	1721	3289
Query	601	99.9%	99.9%	1721	3289
Db	636	99.9%	99.9%	1721	3289
Query	661	99.9%	99.9%	1721	3289
Db	696	99.9%	99.9%	1721	3289
Query	721	99.9%	99.9%	1721	3289
Db	756	99.9%	99.9%	1721	3289

RESULT 9
 ID W07585 standard: protein; 812 AA.
 AC W07585;
 DT 24-JUN-1997 (first entry)
 DE Murine plasminogen sequence.
 KW angiotatin; plasminogen; kringles; angiogenesis; cancer; arthritis;
 macular degeneration; diabetic retinopathy.
 OS Mus musculus.
 PN W09635774-A2.
 PD 14-NOV-1996.
 PF 26-APR-1996; U05856.
 PR 26-APR-1996; US-429743.
 PR 22-FEB-1996; US-605598.
 PR 08-MAR-1996; US-612788.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PI Cao Y, Folkman MJ, Lin J, O'Reilly MS, Slin KL.
 DR WPI; 96-518662/51.
 PT Use of angiotatin fragments or aggregates - for inhibiting
 PT endothelial cell proliferation and treating angiogenesis-mediated
 PT diseases, e.g. cancer, arthritis or diabetic retinopathy
 PS Disclosure: Fig 1: 203pp; English.
 CC The invention relates to new methods and compositions for
 CC inhibiting endothelial cell proliferation, using as active component
 CC an angiotatin fragment, a combination of angiotatin fragments, or
 CC aggregate angiotatin. The fragment is preferably derived from murine,
 CC human, Rhesus, porcine or bovine plasminogen and is a kringles 1,
 CC kringles 2, kringles 3, kringles 2-3, kringles 1-3, kringles 1-2, kringles
 CC 1-4 or kringles 1-4BKUS protein. The aggregate angiotatin has a Mol. Wt.
 CC of 45-65 kD and is derived from a plasminogen fragment beginning at
 CC approximately amino acid number 98 of murine, human, Rhesus, porcine or
 CC bovine plasminogen. The active component can be used for treating
 CC angiogenesis-mediated diseases such as cancer, arthritis, macular
 CC degeneration and diabetic retinopathy. It can also be used to develop
 CC antibodies for use in diagnosis, detection and therapy.
 CC The present sequence is the whole murine plasminogen sequence from
 CC which fragments or aggregates may be derived for use in the invention.
 SQ Sequence 812 AA;

Query Match 17.3%; Score 634; DB 1; Length 812;
 Best Local Similarity 41.4%; Pred. No. 1,33e-50; Indels 13; Gaps 12;
 Matches 111; Conservative 51; Mismatches 93;

DB 553 DYCDIPICASSSEFCGKPEVPEKCGRVGCVANPMSMPQISLRTFTGQHFCCGT 612
 QY 228 DACSSKAVVSLRCLACG-VNLNSR-OSRIVGGSALPGAMPQVSLAHVQ-NV-HYCGGS 283
 DB 613 LIAPEWVLTAAHCLKS-SRP-EF--YKVIL-GAHEEYIRGLDVOE-ISVAKLILEPNNR 666
 QY 284 IITPEWVLTAAHCEKELNPNMWTAFAGILROSFMEYGAQYVEKYSHPNDKTKNN 343
 DB 667 DIALKLSRPATITDKYIPACLPSPNMTVADRTICYTGGEGTGGAGR-LKEADLPV 725
 QY 344 DIALMKLQKPLTFNDIVKPVCLPBPGLMLOPEOLCWSIGMGATEKGTSEVLAAKVL 403
 DB 726 IEHKVVCN-RVEYLNRRKSTELCAGLAGVDSQSGSGPLVCFEEDKYLIGSTSMGL 784
 QY 404 IETGRCHSRIVY-DNLITTPMICAGFLQGVNDSCQSDSGPLVTSKNNIMLIDTSMGS 462
 DB 785 GCARPKNPGYVVRVREVDIEMERNNN 812
 QY 463 GCAKAYRPGYGVNMYVTDMITYRQMRAD 490

RESULT 10
 ID W94036 standard: protein; 812 AA.
 AC W94036;
 DT 12-APR-1999 (first entry)
 DE Murine plasminogen.
 KW Plasminogen; kringles; endothelial; angiogenesis; tumour; leukemia;
 rheumatoid arthritis; psoriasis; ocular angiogenic disease; ulcer;
 gene therapy; birth control; Crohn's disease; angiotatin.
 OS Mus sp.

FH Key Location/Qualifiers
 FT Domain 103..181
 FT /note="Kringles 1"
 FT 185..262
 FT /note="Kringles 2"
 FT 275..352
 FT /note="Kringles 3"
 FT 377..454
 FT /note="Kringles 4"
 FT 185..352
 FT /note="Kringles 2-3"
 FT 103..352
 FT /note="Kringles 1-3"
 FT 103..262
 FT /note="Kringles 1-2"
 FT 103..454
 FT /note="Kringles 1-4"
 FT 93..470
 FT /note="Kringles 1-4BKUS"

PN W09854217-A1.
 PD 03-DEC-1998.
 PF 29-MAY-1998; U10979.
 PR 30-MAY-1997; US-866735.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PI Folkman MJ, O'Reilly MS;
 DR WPI; 99-059809/05.
 PT Use of plasminogen fragments - having an amino acid sequence similar
 PT to the kringles 1-5 region, for inhibiting endothelial cell
 PT proliferation and angiogenesis
 PS Examples: Fig 1A-B: 105pp; English.
 CC The invention relates to inhibition of endothelial cell proliferation
 CC that comprises administering to an endothelial cell a plasminogen
 CC fragment having an amino acid sequence similar to the kringles 1-5 region
 CC of a plasminogen molecule. The plasminogen fragments can be derived from
 CC murine, human, Rhesus, porcine or bovine plasminogens. The plasminogen
 CC fragments can be used for modulating angiogenesis and treating angiogenic
 CC mediated disease e.g. solid tumours, blood born tumours (leukaemia),
 CC tumour metastasis, benign tumours (e.g. haemangiomas, acoustic neuromas,
 CC neurofibromas, trachomas and pyogenic granulomas), rheumatoid arthritis,
 CC psoriasis, ocular angiogenic diseases (e.g. diabetic retinopathy,
 CC retinopathy of prematurity, macular degeneration, corneal graft
 CC rejection, neovascular glaucoma, retrolental fibroplasia, rubeosis),
 CC Osler-Webber Syndrome, myocardial angiogenesis, haemophilic joints,
 CC plaque neovascularisation, telangiectasia, angiodioma, wound
 CC granulation. The fragments can also be used for treating excessive or
 CC abnormal stimulation of endothelial cells (intestinal adhesions), Crohn's
 CC disease, atherosclerosis, scleroderma and hypertrophic scars; as a birth
 CC control agent by preventing vascularisation required for embryo
 CC implantation; and in the treatment of diseases that have angiogenesis as
 CC a pathological consequence such as cat scratch disease (Rochelle minia
 CC guttosa) and ulcers (H. pylori). The nucleotide sequences encoding the
 CC plasminogen fragments can also be used for gene therapy. The products can
 CC be used for the production of antibodies and in detection and diagnosis.
 CC The present sequence represents a murine plasminogen sequence. The
 CC kringles regions that can be used in the invention are indicated in the
 CC features.
 SQ Sequence 812 AA;

Query Match 17.3%; Score 634; DB 1; Length 812;
 Best Local Similarity 41.4%; Pred. No. 1,33e-50; Indels 13; Gaps 12;
 Matches 111; Conservative 51; Mismatches 93;

DB 553 DYCDIPICASSSEFCGKPEVPEKCGRVGCVANPMSMPQISLRTFTGQHFCCGT 612
 QY 228 DACSSKAVVSLRCLACG-VNLNSR-OSRIVGGSALPGAMPQVSLAHVQ-NV-HYCGGS 283
 DB 613 LIAPEWVLTAAHCLKS-SRP-EF--YKVIL-GAHEEYIRGLDVOE-ISVAKLILEPNNR 666
 QY 284 IITPEWVLTAAHCEKELNPNMWTAFAGILROSFMEYGAQYVEKYSHPNDKTKNN 343
 DB 667 DIALKLSRPATITDKYIPACLPSPNMTVADRTICYTGGEGTGGAGR-LKEADLPV 725
 QY 344 DIALMKLQKPLTFNDIVKPVCLPBPGLMLOPEOLCWSIGMGATEKGTSEVLAAKVL 403


```
Db 624 CTGCAATGCGCGCTGACTCTATGGAACAGATCAAGCCAGAGATGTCTGTGCTG6CTA 683
Qy 1398 cctgcaggaggaaacgltcgtatcttgcagggtgcagatgcaggccctcgtg----- 1449
Db 684 CCGCGAGGGTGGCATTGATGCCAGGCGCGACAGCGGTGTCCCTTGTGTGTGAGGA 743
Qy 1449 ---caactgcgaagaacaatatactgtgtgcgtatagaggatatacaagctgagggttcgtgctg 1505
Db 744 CAGCATCTCTCGAGCCGACAGTTGGCGGCTGTGTGTCATTGTGAGTTGGGCACTGGCTG 803
Qy 1506 tgcgaagccttacagaccagaggtgcacggagatgcaggtatcgcaggaactgcattta 1565
Db 804 TGGCCTGGGCCAGAGACGAGGCGCTTACACCAAGTCAGTGACTTCGGGAGTGAATCTT 863
Qy 1566 tcgacaatatgcagggcagagcgtctaatcca 1594
Db 864 CCAGGCCATTAAGACTACACTCCGAGGCA 892
```

Search completed: April 22, 2000, 06:03:00
Job time: 2649 sec